CRYSTALLOGRAPHIC STUDIES ON MOLECULES OF BIOLOGICAL IMPORTANCE

Peter James Gilchrist

Submitted in satisfaction of the requirements

for the degree of PhD in the

University of Edinburgh

Year of presentation 1992

EX BIBL. UNIV. EDINBURGEN.

Acknowledgements

I gratefully acknowledge the financial support provided by Sandoz AG during my three years at the department of Biochemistry and the nine months spent in Basle, Switzerland.

I would like to express my gratitude to the following friends and colleagues who have helped me in the work of this thesis:

Dr M. D. Walkinshaw my industrial supervisor, and Dr. L. Sawyer my academic supervisor, for their patient supervision and advice throughout the course of this work.

Dr P. Taylor for all his help in coming to terms with computing.

Mr D. Penman for assistance with the electron microscopy.

To Alan, Paul, Mary, Joao, Elspeth, Kevin, Jeremy, Tim, Stella, Fred and all the other people who have made my time in Room 302 very enjoyable.

Finally, I must thank my family and friends for their support and encouragement.

This thesis is dedicated to three women, Niamh, my mother and my grandmother. Thank you.

Declaration

I hereby declare that this thesis was composed by me, that the work of which this is a record was done by me, except where stated in the thesis. This work has not been accepted elsewhere in any previous application for a degree. All of the sources have been acknowledged.

List of contents

List of figures

List of Tables

List of Plates

List of Abbreviations

$A^{1\%}_{1cm}$	absorbance of 1% solution (w/v) in a 1cm path length
A_{280nm}	absorbance at 280nm
Å	angstroms
a*,b*,c*	unit cell dimensions of reciprocal space
AD	Alzheimers disease
B	temperature factor
CD	circular dichroism
CG	conjugate gradient
CGRP	calcitonin gene-related peptide
CM-52	carboxymethyl cellulose
C.O.G.	centre of gravity
DE-52	diethylaminoethyl cellulose
DMF	dimethyl-formamide
DMPG	dimyristoyl-phosphatidylglycerol
DMSO	dimethyl-sulphoxide
DNA	deoxyribonucleic acid
$E_{SBOUNDS}$	energy term for mean field boundaries of solvent
E_{PVDW}	energy term for van der Waals packing interactions
E_{PELEC}	energy term for electrostatic packing interactions
FFT	fast fourier transform
$\mid F_{hkl} \mid$	structure factor amplitude

$\| F_o \|$	observed structure factor amplitude
$\| F_c \|$	calculated structure factor amplitude
FPLC	fast performance liquid chromatography
f	atomic scattering factor
HPLC	high performance liquid chromatography
5HT	5-hydroxytyptamine
h,k,l	Miller indices
i	the square root of -1
K	temperature in degrees Kelvin
k_B	Boltzmann constant 1.38×10^{-23} JK^{-1}
K_{Ca}	calcium binding constant
\	wavelength
L	Lorentz factor
M	molar
MPD	2-methyl-2,4-pentanediol
M_r	relative molecular mass
NMR	nuclear magnetic resonance
3D-NMR	three-dimensional nuclear magnetic resonance
p	polarisation correction
ps	pico-second
PEG	polyethylene glycol
PLA-2	phospholipase A2
PMSF	phenylmethylsulphonylfluoride
R	conventional residual $\Sigma_{hkl}(\| F_o \| - \| F_c \|)/\Sigma_{hkl}(\| F_o \|)$
rms	root mean square
SA	Simulated annealing
SD	steepest descent
SDS-PAGE	sodium dodecylyl-sulphate polyacrylamide gel

	electrophoresis
sCT	salmon calcitonin
SEM	scanning electron microscopy
TEM	transmission electron microscopy
TFE	tri-fluoroethanol
Θ	angle of incidence to the lattice plane
w/v	weight per volume
v/v	volume per volume

Chapter 1 Introduction

Introduction

Drugs frequently exert their action by the specific interaction with some target protein. Only as a consequence of binding do the drugs produce their biological effect. A complete understanding of biological interactions can only be achieved at the molecular level. This requires the use of techniques such as x-ray crystallography or 3-D NMR. Armed with a structural knowledge of these interactions it is possible to exploit this information making a number of things possible such as the development of novel compounds which interact with more strength and the development of compounds that bind without inducing biological effect. In this thesis, the determination of the crystal structure of equine phospholipase A2 as outlined in Chapters 5 and 6 was performed with this ultimate aim in mind. It is hoped that in the determination of this structure it would be possible to study the interactions of this enzyme with existing inhibitors and suggest novel or improved versions of these.

There are a number of approaches possible in the search for this structural information. The approach taken in any investigation is usually dictated by the material available to the investigator. The best approach to the study of these interactions, but often the most difficult to produce, is to determine the crystal structure of the receptor and the drug in a complex. In this thesis an existing inhibitor of equine PLA-2 was used in the attempted structural determination of the enzyme-complex as described in Chapter 4 and 5. If a complex is not available but rather the three dimensional structure of both molecules are available, it is possible to "dock" the two molecules together using molecular graphics and infer the

nature of their interactions. However if the structure of the receptor is not available this is more difficult. Models for the receptor can be obtained however if a homologous protein whose structure has been determined is available. If no such homologous receptor structure is available it is still possible to get information, such as the shape of the binding site and spatial disposition of functional groups, by the comparison of the crystallographic structures of compounds which are known to bind to the receptor. This is the rationale behind the investigation of the crystal structure of an adrenoreceptor antagonist and it's comparison to similar compounds in Chapter 2. It is hoped that structural features of these small ligands can be identified as being important to their interaction with their receptor. The attempted crystallisation of calcitonin and calcitonin analogues in Chapter 4 is similarly with a view to understanding important structural features involved in the interaction with its receptor. In the absence of any material, providing the sequence is available, it is still possible to produce some model of the receptor using structure prediction methods and modelling techniques. These methods are unlikely to produce useful results since our understanding of protein architecture is not sufficiently advanced and the resulting model will very likely be wrong.

1.1.1 Phospholipase A2 as a target for drug design

In a great many disease states, pain and disability are caused by inflammation. The inflammatory process is mediated within the body by the production of eicosanoids. The eicosanoids are a family of molecules with diverse biological activities, but which are produced from the same starting material. This starting compound is called

arachidonic acid (5,8,11,14-icosatetraenoic acid). Eicosanoid is the term introduced by Corey *et al.* (1980) to cover all products of arachidonate metabolism, whether by the cyclo-oxygenase pathway (prostaglandins, endoperoxides, thromboxanes, etc.) or the lipoxygenase pathway (hydroxyicosatetraenoic acids, leukotrienes, etc.). At present, with the exception of steroidal compounds, anti-inflammatory compounds have been developed to inhibit only the cyclo-oxygenase pathway. The most well known of these compounds, aspirin, acts directly on the cyclooxygenase enzyme causing inhibition. Eicosanoid production can therefore only be reduced with these inhibitors and not obliterated, since alternate pathways remain active. A more suitable target for the design of anti-inflammatory compounds is another enzyme involved in the pathway of eicosanoid production, phospholipase A2, which releases the starting material, arachidonic acid, from phospholipids. Inhibition of this enzyme would severely reduce the levels of arachidonate and consequently all eicosanoids. It has been established that this enzyme plays an important role in the inflammatory processes associated with rheumatoid arthritis (Pruzanski & Vadas, 1988), osteoarthritis (Vignon *et al.*, 1989), adult respiratory distress syndrome (Vadas & Pruzanski, 1986), and gout (Bomolaski *et al.*, 1990).

A wide variety of compounds has been shown to inhibit PLA-2, of which the best characterised inhibitor is p-bromophenacylbromine, an irreversible inhibitor that specifically modifies the active site histidine residue (Roberts *et al.*, 1977). Other covalent inhibitors are the natural product monoalide (Glaser & Jacobs, 1986) and it's synthetic analogue monoalogue (Reynolds *et al.*, 1988) which have been shown to modify covalently lysine residues. There are also a number of

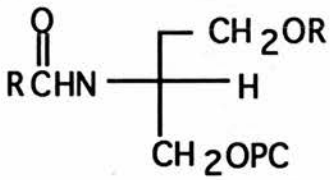
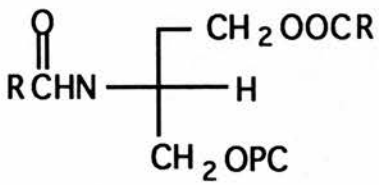
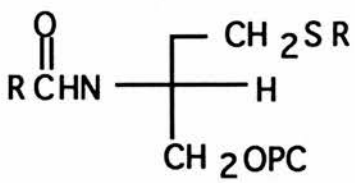
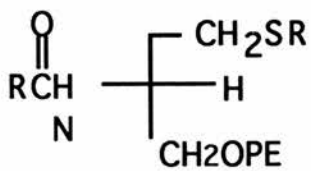
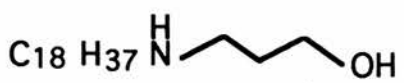
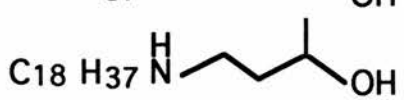
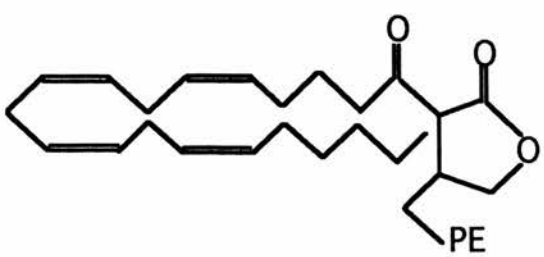
protein inhibitors of PLA-2. These produce their inhibitory effect mainly through substrate depletion, consequently their physiological significance is still in doubt (Davidson *et al.*, 1987). These proteins are very similar in both sequence and mechanism of inhibition and have been classed together as the annexins and include the lipocortins and calpactins (Klee, 1988). These are Ca^{2+} dependent membrane binding proteins which can be phosphorylated by growth factor receptor kinases. The anti-inflammatory action of glucocorticoids is known to be through the inhibition of phospholipase A2 activity. The enzyme is not inhibited directly, rather glucocorticoids are believed to induce the synthesis of these annexins (Flower, 1989).

The vast majority of anti-PLA-2 compounds studied so far are non-hydrolysable phospholipid analogues. These have provided valuable information on the structural features of the phospholipid which are important for binding. Since the hydrolysis of the phospholipids is a stereospecific reaction, the D-isomer of the phospholipids is an example of such a non-hydrolysable substrate. It has binding characteristics similar to the natural substrate and is therefore of no potential use as an inhibitor. A variety of these phospholipid analogue inhibitors and their inhibitory activity are shown in table 1.1.

There are difficulties in comparing the efficacy of these inhibitors since it has been shown that PLA-2 activity is crucially dependent not only on the nature of the substrate (de Haas *et al.*, 1971), but also on the physical state of the lipid (Lichtenberg *et al.*, 1986). Also, since there are many different methods available for the measurement of enzyme activity, comparisons are not necessarily straightforward (Dennis, 1991). Where more than one method has been used to measure

Table 1.1

Phospholipid analogue inhibitors of PLA-2 activity. Inhibition is quoted as IC_{50} values. These are the concentration of inhibitor required to reduce enzyme activity by 50%. Phosphatidylcholine and phosphatidylethanolamine are represented by PC and PE respectively. R represents the palmitoyl (C16) group.

Name	Structure	IC_{50}
1. Ether Amide PC		38μM
2. Ester Amide PC		156μM
3. Thioether Amide PC		2μM
4. Thioether Amide PE		0.45 μM
5. Long Chain Alkyl Amines		20μM
		10μM
6. 3-Arachidonyl-1,2,3,4-tetrahydro-4[O-Phosphatidyl ethanamino hydroxymethyl] furan-2-one		64 μM

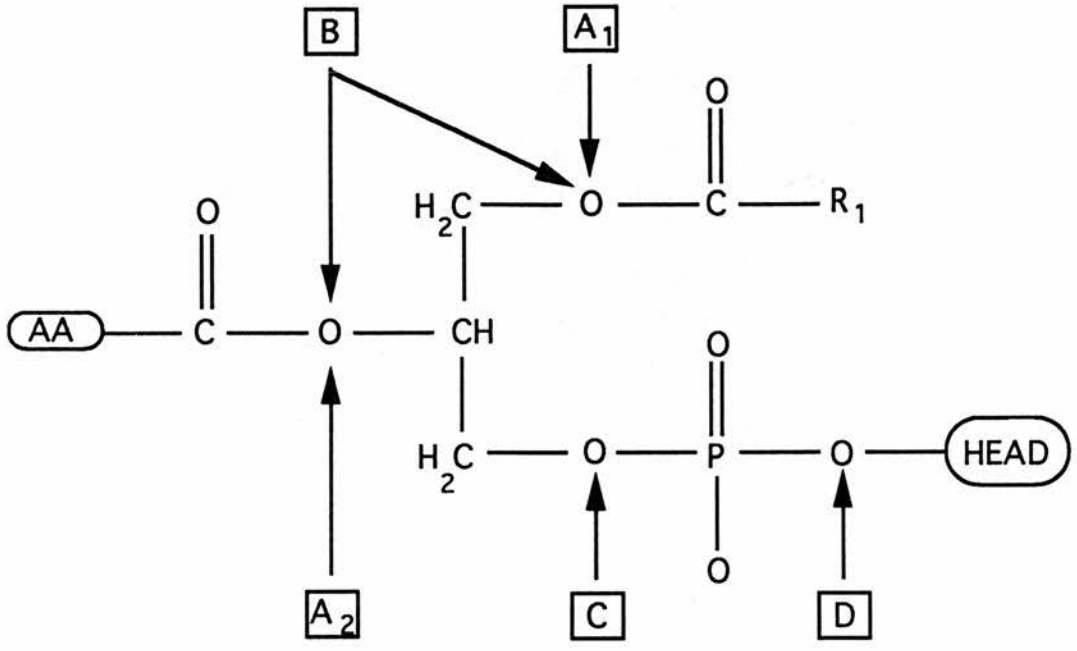
inhibition by a particular inhibitor, quite different values have been obtained (Davidson *et al.*, 1986). Inhibition studies of water-soluble lipolytic enzymes which act on water-insoluble substrates are complicated. Thus lipolysis is an example of heterogeneous catalysis and requires a slightly more complicated kinetic model to interpret inhibition results (Verger & de Haas, 1976). This has frequently not been applied in the interpretation of inhibitor studies.

One of the major problems in designing an inhibitor which is a non-hydrolysable substrate analogue is that the inhibitor must be very tightly bound in order to compete successfully against the natural substrate which is vastly in excess. The other problem is that these compounds are, by their nature, hydrophobic and therefore not water-soluble. As is obvious from the IC_{50} values shown in table 1.1 these types of inhibitors are far from being effective inhibitors for clinical use. Further work using crystallographically determined structures and inhibitor complexes may result in overcoming both of these problems to produce some novel compounds with specific phospholipase A2 inhibitory activity for clinical use.

1.1.2 Biochemistry of phospholipase A2.

The hydrolysis of various ester bonds in phospholipids is catalysed by a number of different phospholipases (figure 1.1). Phospholipase A2 catalyses the hydrolysis of the acyl ester bond at the sn-2-position of 1,2-diacyl-sn-3-phosphoglycerides (L-isomer). This enzymic activity is absolutely dependent on the presence of calcium ions. Since this 2-acyl position is the position at which the bulk of arachidonate in mammalian cells is esterified in the fatty acyl chains of glycerophospholipids, the action of phospholipase A2 liberates

Figure 1.1 Classification of phospholipases according to the site of hydrolysis. The phospholipid is shown here in the sn-3 configuration. R_1 represents an alkyl chain. HEAD represents any of the naturally occuring substituents (such as choline, serine, ethanolamine) or a proton. The arrows indicate the site of attack by the phospholipases.



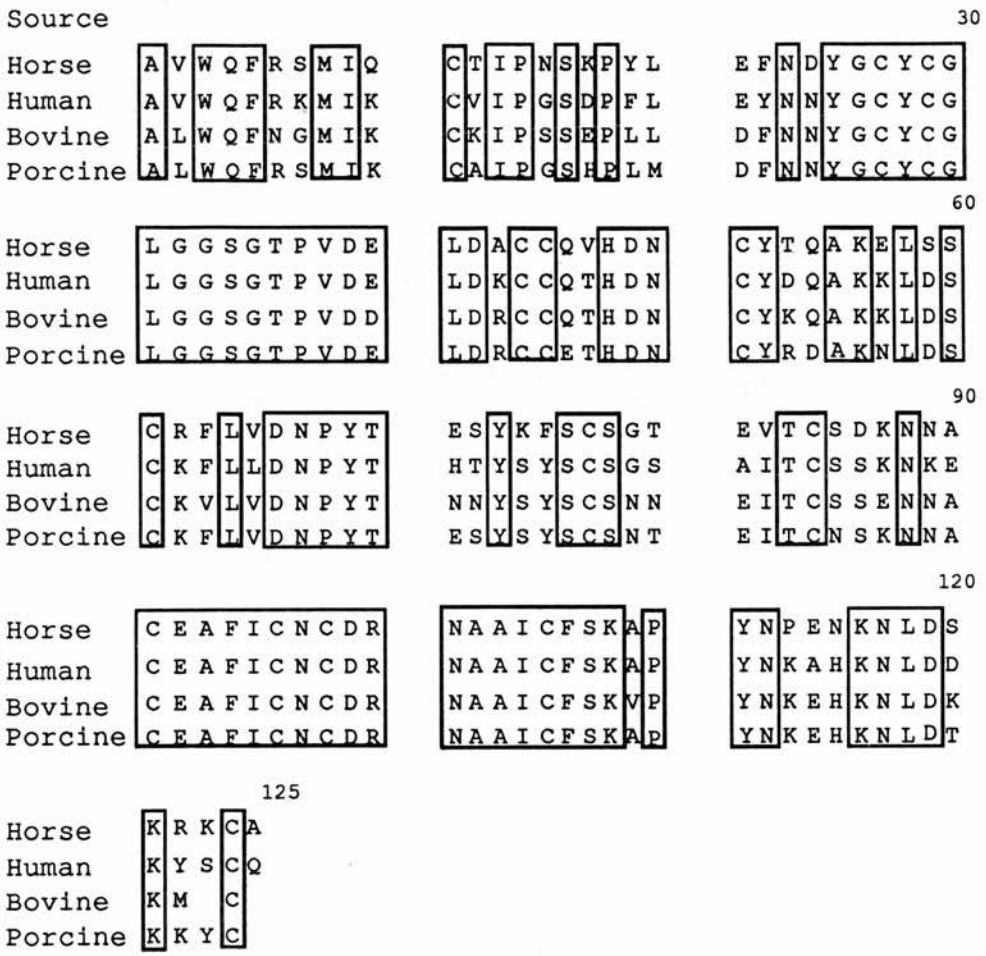
arachidonate. This has led to suggestions that phospholipase A2 may be responsible for controlling free arachidonate levels (Vogt *et al.*, 1966; van den Bosch, 1980). However, it is far from clear that these phospholipases are arachidonyl-specific. Phospholipases A2 (EC 3.1.1.4), both intracellular and extracellular, are widely found in nature (Verheij et al., 1981; van den Bosch 1980). Although intracellular phospholipases will be discussed, the work in this thesis has been carried out on extracellular phospholipases of pancreatic tissue. Other sources of extracellular phospholipases are the venoms of snakes and arthropods. The primary function of the extracellular pancreatic enzymes is in digestion. The snake venom enzymes possess not only neurotoxic but other pharmacological effects such as haemolytic action, anticoagulant properties and myonecrotic properties (Karlsson, 1978; Howard & Gundersen, 1980). Intracellular phospholipases are believed to participate in metabolic regulation of phospholipids in membranes including eicosanoid biosynthesis (van den Bosch 1980), the acylation-deacylation cycle (van den Bosch 1980), and protection of membranes from lipid peroxidation damage (van Kuijk *et al.*, 1987). Recently intracellular phospholipases have been implicated in receptor mediated signal transduction and as part of a G-protein-regulated effector system generating arachidonic acid metabolites that may serve as second messengers (Axelrod *et al.*, 1988). However the intracellular phospholipases have not been as extensively studied structurally, due to their low concentrations, but appear to have the same essential properties as the more abundant pancreatic and venom enzymes (Crews *et al.*, 1981; McGiveney *et al.*, 1981). Consequently the extracellular enzymes have been used as models for the general structure and function of all phospholipases A2. Thus the design of

inhibitors and their action with the extracellular phospholipases are also applicable to the intracellular enzymes.

1.1.3 Sequence properties of phospholipase A2

The extracellular enzymes are single chain polypeptides of between 120-130 residues. The chain contains 14 cysteines connected in 7 disulphide bridges. These bridges are believed to provide the necessary structural stability to allow PLA-2 to act at the hydrophobic/hydrophilic interface that is the membrane surface. This structural stability has also enabled the ready purification of the enzyme from a variety of sources: more than 40 complete amino acid sequences and many more partial structures have been determined (Verheij *et al.*, 1981). Based upon this information, the phospholipases have been separated into two groups (Heinrikson *et al.*, 1977). Group I enzymes are those from mammalian pancreatic juices and the venoms of snake families Elapidae and Hydrophidae. These enzymes possess a disulphide linkage between half cysteines 11 and 69 (numbering from Heinrikson *et al.*, 1977) that is lacking in Group II enzymes. The Group II enzymes are those from the venoms of the snake families Crotalidae and Viperidae and possess a short extension at the COOH terminus that terminates in a half-cysteine linked to a half-cysteine near the active site at position 50 (numbering from Heinrikson *et al.*, 1977). The sequences of some pancreatic enzymes are shown in figure 1.2. Overall, about one quarter of the residues are absolutely conserved; and for another quarter, only conservative changes have occurred (Verheij *et al.*, 1981). It is clear that the phospholipase A2's are homologous proteins which have probably developed from a common ancestor. The venom enzymes unlike the

Figure 1.2 Sequence comparisons of some pancreatic PLA-2's. Residues are denoted by the single letter code and boxed residues indicate identical residues.



pancreatic enzymes, are synthesised in their active forms. The pancreatic enzymes are synthesised as zymogens. The prophospholipase A2 contains an extra heptapeptide at the amino terminus which is cleaved by trypsin to produce the active form. The high sequence similarity of these phospholipases is thought to result in the proteins having a similar fold. Even in the bee-venom phospholipases, which have little sequence homology with other phospholipases, similar "substructures" have been observed and a common catalytic mechanism has been suggested (Scott *et al.*, 1990).

1.1.4 Kinetic properties of phospholipase A2.

The most intriguing aspect of phospholipase kinetics is the observation that although the enzyme can hydrolyse monomeric substrates, the same substrate present as an organised lipid-water interface is degraded at a much higher rate (de Haas *et al.*, 1971). In bulk studies, the influence of a changing lipid-water interface on enzyme activity has been demonstrated using short chain lecithins as substrates (Bonsen *et al.*, 1972). This has been attributed to a preferred conformation of the substrate at the lipid-water interface of the micelle. Changes in monomer phospholipid conformation / orientation as the phospholipid becomes packed in an interface have been observed in ^{1}H and ^{13}C-NMR studies (Roberts *et al.*, 1978; Burns and Roberts, 1980). Kinetic studies on the zymogen of the pancreatic phospholipase indicated an identical activity to the active form with monomeric substrates (Pieterson *et al.*, 1974a). A similar activity for aggregated substrates was not however observed with the zymogen. This led to the idea of an "interface recognition site" (IRS) which is absent in the zymogen.

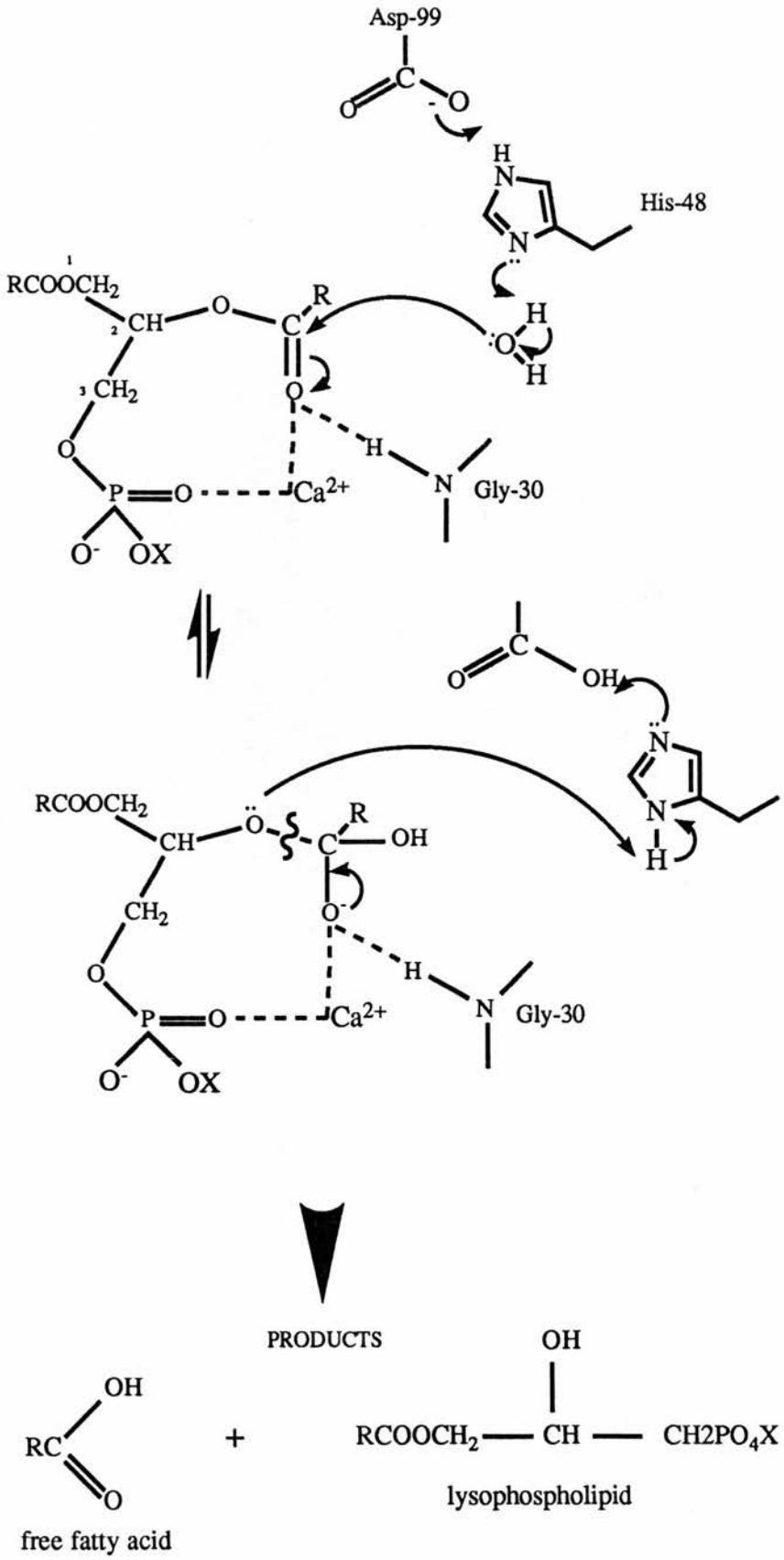
An absolute dependence for Ca^{2+} ions for enzyme activity has been demonstrated in both the zymogen and the active enzyme (Pieterson *et al.*, 1974b). Although the related ions Sr^{2+} and Ba^{2+} can be bound, the enzyme has no activity. However enzyme activity has been observed with transition metal ions, such as Gd^{3+} (Hershberg *et al.*, 1976). The alkaline pH optimum of the enzyme activity (7-8) indicated the involvement of a residue with a pK of 6-7 which was attributed to a histidine (Volwerk *et al.*, 1974). This was subsequently demonstrated to be histidine-48 (numbering of residues as in active enzyme) by chemical modification studies (Volwerk *et al.*, 1974). Since Ca^{2+} protected the modification of this histidine it was concluded that the Ca^{2+} binding site and His-48 were in close proximity at the active site of the enzyme.

1.1.5 Catalytic Mechanism of phospholipase A2.

Although valuable information has been obtained from kinetic analysis, such as the involvement of Ca^{2+} and His-48, the catalytic mechanism has been proposed as a result of the high resolution crystal structure of bovine phospholipase A2 (Dijkstra *et al.*, 1981). The proposed catalytic mechanism (Verheij *et al.*, 1980) is shown in figure 1.3. This model has been supported by a number of other lines of evidence; combined molecular orbital and molecular mechanics calculations (Waszkowycz *et al.*, 1989), the x-ray crystal structure of bee-venom enzyme in a complex with a transition-state analogue (Scott *et al.*, 1990), a similar complex for the cobra-venom enzyme (White *et al.*, 1990) and a substrate analogue complex with a porcine pancreatic phospholipase (Thunnissen *et al.*, 1990b).

The mechanism involves a hydrogen-bonded histidine-aspartate couple

Figure 1.3 Proposed mechanism of catalysis in PLA-2 (Verheij et al., 1980).



in the active site (His-48 and Asp-99 of bovine pancreatic PLA-2), which is similar to other hydrolytic enzymes such as the serine proteases. In PLA-2 there is no serine available locally and it has been proposed that a water molecule in the active site becomes the active nucleophile on proton transfer to histidine as shown in figure 1.3. The attack by water on the C-2 carbonyl of the phospholipid leads to the formation of a tetrahedral oxyanion intermediate, which is stabilised electrostatically by a calcium ion and by hydrogen bonding to the backbone -NH of Gly-30. The oxyanion breaks down, on proton transfer from His-48, to release the C-2 fatty acid.

1.1.6 X-ray Studies on phospholipase A2.

High resolution crystal structures of representatives from Group I and II PLA-2's have been obtained as shown in table 1.2. There are also structures which possess bound inhibitors and have provided information on substrate binding and catalytic mechanism. Comparisons of these structures have subsequently confirmed the earlier sequence analysis predictions regarding the similarities between the group I and II enzymes (Heinrikson *et al.*, 1977). Similar folds have been observed between members of the two different groups. This is most noticeable with respect to the disposition of the active site residues which are highly conserved (Scott *et al.*, 1990).

The pancreatic enzymes are essentially kidney shaped, with dimensions of approximately 22 x 30 x 42 Å. In the bovine structure (Dijkstra *et al.*, 1981a), there is an appreciable content of secondary structure: 50% of the residues are in an α-helical conformation and 10% in an antiparallel 2-stranded ß-structure. The core of the molecule consists of two antiparallel α-helices, each having five full turns as shown in

Table 1.2
Some existing crystal structures of PLA-2. The classification of the enzymes are indicated. Bee venom PLA-2 which does not belong to class I or II is designated by the symbol ?.

Structure	Resolution (A)	Class	Reference
Bovine PLA-2	1.7	I	Dijkstra et al., 1981
Engineered Porcine PLA-2	2.1	I	Thunnissen et al., 1990a
Crotalus Atrox (Dimeric) PLA-2	2.5	II	Brunie et al., 1985
Recombinant human rheumatoid arthritic synovial fluid PLA-2	2.2	II	Wery et al., 1991
Engineered porcine PLA-2 in a complex with a substrate derived inhibitor	2.4	I	Thunnissen et al., 1990b
Cobra venom PLA-2 in a complex with a transition-state analogue	2.0	I	White et al., 1990
Bee venom PLA-2 in a complex with a transition-state analogue	2.0	?	Scott et al., 1990

figure 1.4. The N-terminal region is also in an α-helical conformation with the amino-terminal NH_3+ of residue 1 buried in the interior of the protein. There are two other helices (B and D in fig. 1.4) each of about one turn. Two antiparallel ß-strands are present at the surface of the molecule. The location of the seven disulphide bridges are also shown in fig. 1.4.

The Ca^{2+} ion, essential for enzyme activity, is bound in the active site which is located in a cavity at the molecular surface. It is surrounded by seven oxygen ligands as shown in figure 1.5. Five ligands are at the corners of an octahedron at an average distance of 2.39Å from the Ca^{2+}. The sixth octahedral position is occupied jointly by one of the Asp 49 carboxyl oxygen atoms (OD_1 at 2.69Å) and a water molecule at 2.66Å from the Ca^{2+}. The main chain oxygen ligands of the Ca^{2+} belong to residues 28, 30, and 32. They are part of the calcium binding loop that runs from residues 25-42. The calcium binding loop is stabilised by two disulphide bonds, Cys 27-123 and Cys 29-45. This stabilization of the calcium binding loop guarantees a precise geometry of the Ca^{2+} with respect to the active site residues.

Despite these high resolution structures, an explanation for the increased activity on aggregated substrates has not been proposed. A few features of the face which is thought to interact with the membrane have been identified. The area around the active site which consists of essentially hydrophobic residues is believed to be responsible for the interaction of the enzyme with the fatty-acyl chains during catalysis. These residues are not absolutely conserved but their character is usually maintained. Contact with the membrane has been attributed to two areas at both ends of the molecule which contains both hydrophobic and polar residues. There is always an

Figure 1.4 Diagramatic representation of bovine PLA-2 (from Dijkstra et al., 1981).

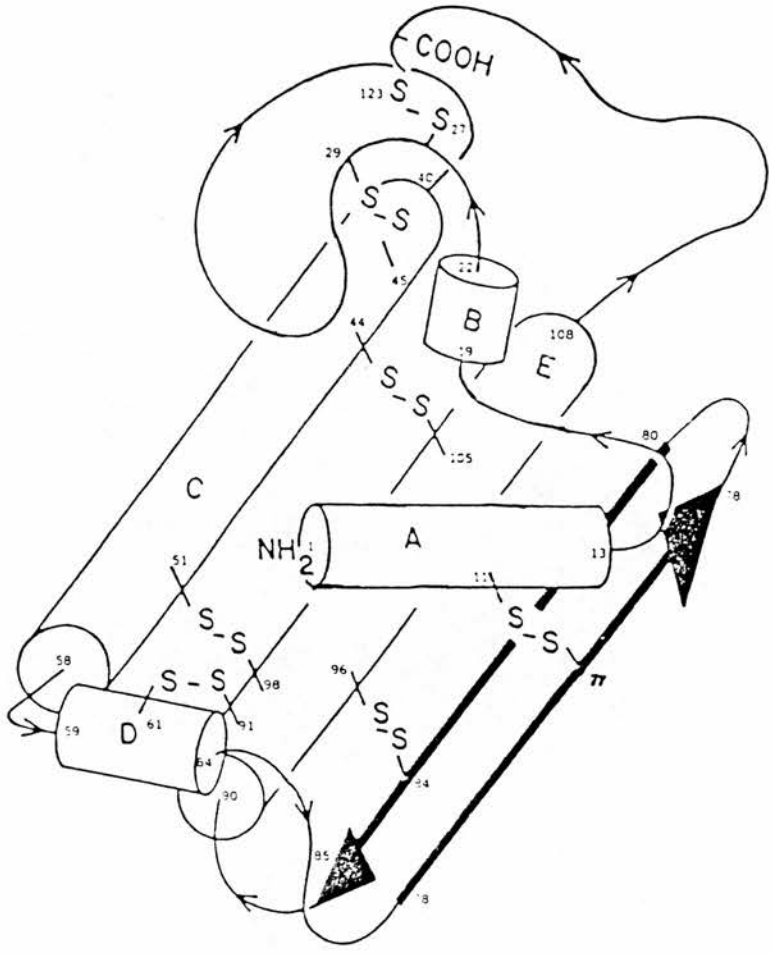
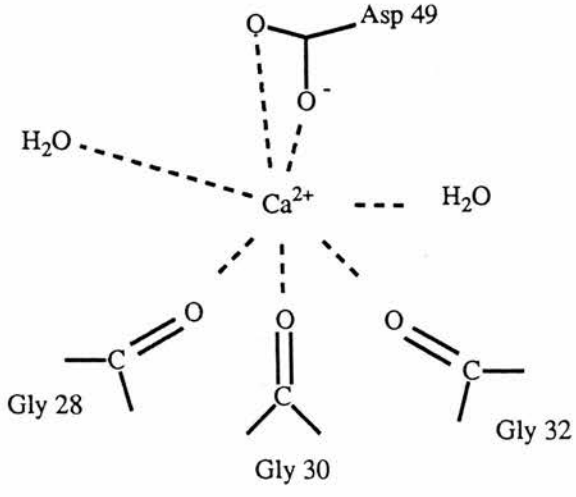


Figure 1.5 Diagramatic representation of the Ca^{2+} binding site of bovine phospholipase A_2 (Dijkstra et al., 1981).



excess of basic over acidic residues in these areas, suggesting salt bridge interactions with phosphate groups of the membrane phospholipids. This preferred orientation has been proposed as partly responsible for the high activity towards aggregated substrates (Dijkstra *et al.*, 1981b).

1.2 Calcitonin and related peptides as a target for drug design.

Small polypeptide hormones are ideal candidates for rational drug design. The small sizes of these polypeptides are easily within the realms of synthetic synthesis. Thus peptide hormone analogues can be obtained in sufficient quantities and purity to be tested with a view to determining quantitative structure-activity relationships. The polypeptide hormone calcitonin is a suitable peptide for this type of investigation. Since the receptor for calcitonin has not been isolated, only studies of active and inactive analogues can be of use to determine what groups or structural features are important in the binding of calcitonin with it's receptor. Additionally, since calcitonin has sequence and structural similarities to both calcitonin gene-related peptide (CGRP) and amylin, it is possible that there are also some general similarities in the nature of the interaction between these peptides and their receptors. The role of an amphipathic α-helix in the interaction of these peptides with their receptors has been suggested (Moe & Kaiser, 1985). The similar nature of the interaction of the peptides with their receptors has already been demonstrated for two of these peptides, calcitonin and CGRP. These have distinct receptors but have the ability to cross react at each others receptors (Goltzman & Mitchell, 1985). It is therefore possible that there is some general structure-activity relationship common to

all three peptides. However, since the peptides themselves have such widely differing biological activities, it would be highly unlikely that their receptor interactions were too closely related. Structural studies on calcitonins and analogues of calcitonins may therefore provide some valuable information about the interaction of this peptide with it's receptor and possibly lead to the development of more active forms of the hormone.

1.2.1 Biochemistry of calcitonin.

Calcitonin is a polypeptide hormone consisting of 32 amino acids, synthesised and secreted by the parafollicular, or C-cells, of the thyroid in mammals. In submammalian vertebrates such as fish, birds and reptiles, the secretion is from a discrete gland, called the ultimobranchial body. The principal pharmacological action of calcitonin in mammalian species is inhibition of bone resorption (O'Riordan & Aurbach, 1968). Treatment with calcitonin over prolonged periods also causes a decline in osteoclast numbers. This latter effect is of great importance in the therapy of Paget's disease and, more recently, of osteoporosis. However, the relevant receptor characteristics for this reduction in osteoclast numbers is unknown. Calcitonin also promotes urinary clearance of calcium and phosphate (Cochran *et al.*, 1970). Calcitonins have been chemically synthesised from several species, porcine (pCT), human (hCT), and salmon (sCT) (Guttmann, 1981) and are available for therapeutic use. Eel calcitonin (eCT) is also available in the market but, unlike other CT's, is produced as the amino-suberic acid derivative (ASU-eCT), since it's natural form is highly unstable (Morikawa *et al.*, 1976). The hypocalcemizing effect of CT, usually measured by a rat bioassay,

differs widely between the various hormone species. The estimated potency of mammalian CT's is between 100-400 international units (IU)/mg peptide, whereas nonmammalian CT's show higher potencies ranging between 2000-6000 IU/mg peptide (Guttmann, 1981). One IU corresponds to 1/100 of the amount of CT needed to induce a 10% reduction of blood calcium 1 hour after intravenous injection into a young fasting 150g rat. Also, the hypocalcemizing effect is more prolonged for nonmammalian CT's (Galente *et al.*, 1976;Genari *et al.*, 1981). Calcitonin receptor sites have been identified on whole bone sections (Warshawsky *et al.*, 1980) and on 'osteoclast rich' cultures (Luben *et al.*, 1976). Receptors have also been identified in non-bone cells, such as kidney (Marx *et al.*, 1972), brain (Tschopp *et al.*, 1985), testes (Chausmer *et al.*, 1982), lymphoid cell lines (Marx *et al.*, 1974) and human cancer cell lines (Findlay *et al.*, 1982). Binding of Calcitonin to it's receptor is associated with stimulation of adenylate cyclase (Nicholson *et al.*, 1986).

1.2.2 Biochemistry of calcitonin gene-related peptide

With the advent of recombinant DNA technology the existence of a related peptide, calcitonin gene-related peptide (CGRP) was predicted in rats (Rosenfeld *et al.*, 1983) and it's existence was subsequently established in man (Morris *et al.*, 1984). This gene, commonly referred to as the α-gene, encodes α-CGRP. More recently, a second calcitonin gene has been described (Hoppener *et al.*, 1986), commonly referred to as the ß-gene. This gene encodes another CGRP molecule (Jonas *et al.*, 1985) and a second calcitonin-like peptide ß-CGRP (Alevizaki *et al.*, 1986), although the latter peptide is not expressed either in the rat or in humans. α-CGRP has subsequently been found to have a widespread

distribution in the body, being particularly prevalent in the central nervous system and peripherally in nerves of the cardiovascular system (Gibson *et al.*, 1984). Calcitonin and CGRP are obtained from the same gene by a process of alternate splicing at the mRNA level (Breimar *et al.*, 1988). The CGRP sequence consists of 37 amino acids with a disulphide bridge between positions 2 and 7 and a C-terminal phenylalanine amide residue. There is about 30% sequence identity between the CGRP peptides and salmon calcitonin. CGRP has been shown to have potent cardioexcitatory (Brain *et al.*, 1986) and vasodilatory (Tippins *et al.*, 1986) effects and has also been shown to lower plasma calcium (Tippins *et al.*, 1984). More recently it has been demonstrated to be a potent inhibitor of glycogen synthesis in rat skeletal muscle *in vitro* (Leighton & Cooper, 1988). Discrete receptor sites for CGRP have been identified in the nervous system and peripheral tissues. It has also been shown that sCT and CGRP can cross react at each others receptors. CGRP has been shown to be 1000-fold less potent than human calcitonin on interaction with CT receptors in an osteoclastic bone resorption assay (Zaidi *et al.*, 1987). Also, in contrast to CT receptors, neural CGRP receptors are not linked to adenylate cyclase (Goltzman & Mitchell, 1985). The cardiovascular CGRP receptors are slightly different to the neural receptors in that they do not cross react with sCT and these are linked to adenylate cyclase (Kubota *et al.*, 1985).

1.2.3 Biochemistry of amylin

Amylin is a 37-amino-acid peptide which is the major component of amyloid deposits in the islets of Langerhans (Cooper *et al.*, 1987). These amyloids are a common feature in human type 2 diabetic subjects

and have been described in a variety of diabetic animals such as the domestic cat (Johnson & Stevens, 1973) and the monkey *Macaca niagra* (Howard, 1986). Amyloid is the term used for a group of pathological extracellular proteinaceous substances that share common morphological characteristics and appear in a diverse number of clinical settings as shown in table 1.3. The main constituents of amyloid are protein fibrils whose protein subunits vary in the different forms of the disease and the chemical nature of which form the basis of classification. Amyloid may be localised or systemic in distribution. Amyloid is generally identified by it's green birefringence after staining with Congo Red (Glenner, 1980).

Amylin has 46% identity to the rat and human calcitonin gene-related peptides. Like CGRP, the peptide contains a disulphide linkage between amino acids at positions 2 and 7 and possesses a C-terminal amide (in this case, Tyr-NH_2). Like α-CGRP, human pancreatic amylin has been shown *in vitro* to inhibit both basal and insulin stimulated rates of glycogen synthesis. It has been suggested that amylin may be a newly recognised hormone with an opposing role to insulin in the homeostasis of glucose metabolism (Leighton & Cooper, 1988). No receptor has yet been identified for amylin and there is no evidence that amylin can bind to either calcitonin or CGRP receptors.

1.3 Sequence comparisons of calcitonin and related peptides.

Sequence comparisons can be very useful between related molecules since residues that are important for biological activity and structure are highly conserved. Thus it is possible to identify conserved regions of molecules and suggest possible roles for these regions. This is particularily useful information in the design of new

Table 1.3 Clinical amyloidosis syndromes. Adapted from Pepys, (1988).

Distribution of deposits	Clinical syndromes	Fibril proteins and precursors
Systemic amyloidosis	Associated with immunocyte dyscrasia:myeloma, monoclonal gammopathy, ocult dyscrasia	AL fibrils derived from monoclonal immunoglobulin light chains
	Associated with chronic active diseases: reactive systemic amyloidosis	AA fibrils derived from serum amyloid A protein (SAA)
	Hereditary symptoms:	
	Predominantly neuropathic forms (several types)	Prealbumin fibrils, genetic variant of plasma prealbumin
	Predominantly nephrotic forms include that associated with familial Mediterranean fever (autosomal recessive), typical AA amyloidosis	AA fibrils derived from serum amyloid A protein (SAA)
	Predominantly cardiomyopathic forms include	
	Senile systemic amyloidosis (formerly called senile cardiac amyloidosis	Prealbumin fibrils, genetic variant of plasma prealbumin
	Associated with chronic haemodialysis	β2-microglobulin derived from high plasma levels
	Senile amyloidosis (heart, brain, joints, seminal vesicles)	Atrial naturetic peptide related fibrils in isolated atrial amyloid, otherwise not known; different in different organs
Localised amyloidosis	Cerebral amyloid angiopathy and cortical plaques in Alzheimer's disease, senile dementia, and Down's syndrome	β-protein fibrils, precursor encoded on chromsome 21
	Periarticular, bony, and renal amyloid in chronic haemodialysis patients	β2-microglobulin derived from high plasma levels
	Endocrine amyloidosis	Precalcitonin-related fibrils in medullary carcinoma of thyroid; calcitonin gene-realted peptide fibrils in islets of Langerhans; ? other peptide hormones
	Isolated massive nodular deposits (skin, lung, urogenital tract)	AL fibrils derived from monoclonal immunoglobulin light chains
	Primary localised cutaneous amyloid (macular, papular)	? Keratin-derived
	Hereditary syndromes:	
	Hereditary cerebral haemorrhage with amyloidosis	
	Icelandic type	Cystatin C (γ-trace) fibrils, Glu 58 genetic variant of cystatin C
	Dutch type	β-protein
	Cutaneous deposits (bulllous, papular, poikilodermal)	Not known
	Ocular deposits (comeal, conjunctival)	Not known

molecules.

1.3.1 Calcitonin

Several mammalian and submammalian calcitonins have been isolated, sequenced and synthesised: porcine (Baghdiantz *et al.*, 1964); human (Riniker *et al.*, 1968); bovine (Brewer & Ronan, 1969); ovine (Potts *et al.*, 1970); salmon (O'Dor *et al.*, 1968); eel (Otani *et al.*, 1976); rat (Raulias *et al.*, 1976). The sequences of calcitonin from seven species are given in Figure 1.6. These display considerable differences in amino acid composition. Salmon has three forms of calcitonin, the major component sCT I (from here on referred to as sCT) and two minor components sCT II and sCT III as shown in Figure 1.6. For all the calcitonins, the sequence is closely conserved through the first nine residues. There is considerable variation in the remaining positions particularly in the middle of the molecule. However, much of the variation is more apparent than real in that the hydrophobic or hydrophilic properties of the positions are retained.

1.3.2 Calcitonin gene-related peptide

Only the primary structure of human α-CGRP has been determined unequivocally by fast atom bombardment mass spectrometry (Morris *et al.*, 1984). The sequence of the rat α- and ß-CGRP and the human ß-CGRP have been predicted from the nucleotide sequence (Rosenfeld *et al.*, 1983; Hoppener *et al.*, 1985; Steenbergh *et al.*, 1985). These known sequences are highly conserved, as shown in Figure 1.7. Human α-CGRP differs from the ß-form by three amino acids (positions 3, 22 and 25). The predicted rat α-CGRP differs from the human sequence by four amino acids (positions 1, 3, 25 and 35; three of these involve charge

Figure 1.6 Sequence alignment of calcitonins from different species. Residues common to sCT I are enclosed in shaded boxes. The single letter code is used to describe the amino acids. The (NH2) group represents the C-terminal amide.

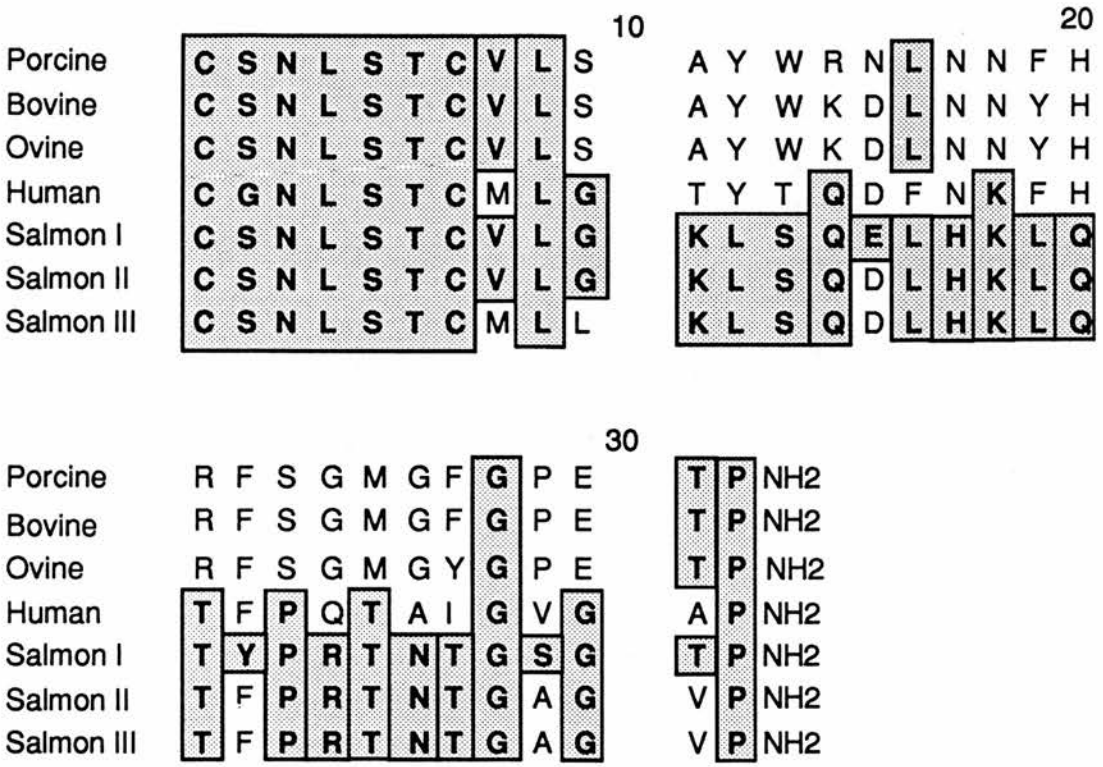
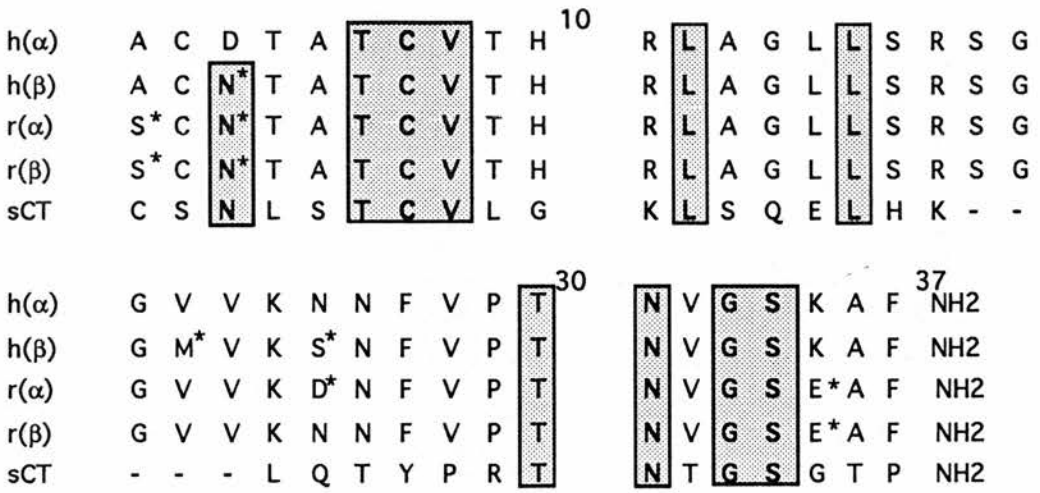


Figure 1.7

Sequence alignment of four CGRP molecules with salmon calcitonin (sCT). A gap of five amino acids, denoted by (-), has been introduced to maximise alignment. Boxed residues show sequence identity to sCT. The asterisks show amino acid substitution differences between human alpha CGRP and the other CGRP molecules. Human CGRP alpha and beta, h(α) and h(β);rat CGRP alpha and beta, r(α) and r(β).



changes). A gap of five amino acids has to be introduced between Lys18 and Leu19 of sCT to maximise sequence alignment with CGRP.

1.3.3 Amylin

Amylin from pancreatic amyloid of human (Clark *et al.*, 1987), cat (Westermark *et al.*, 1987) and rat (Leffert *et al.*, 1989) has been purified and sequenced. The cat sequence has only been partly assigned but indicates similarity to the human form as shown in figure 1.8. The sequences of human and rat differ at only six residues (84% identity) as shown in Figure 1.8. However, three proline residues have been introduced towards the C-terminal region of rat amylin which may have a profound effect on the structure at the C-terminal region. Like CGRP, a gap of five amino acids has to be introduced into sCT to maximise sequence alignment with amylin.

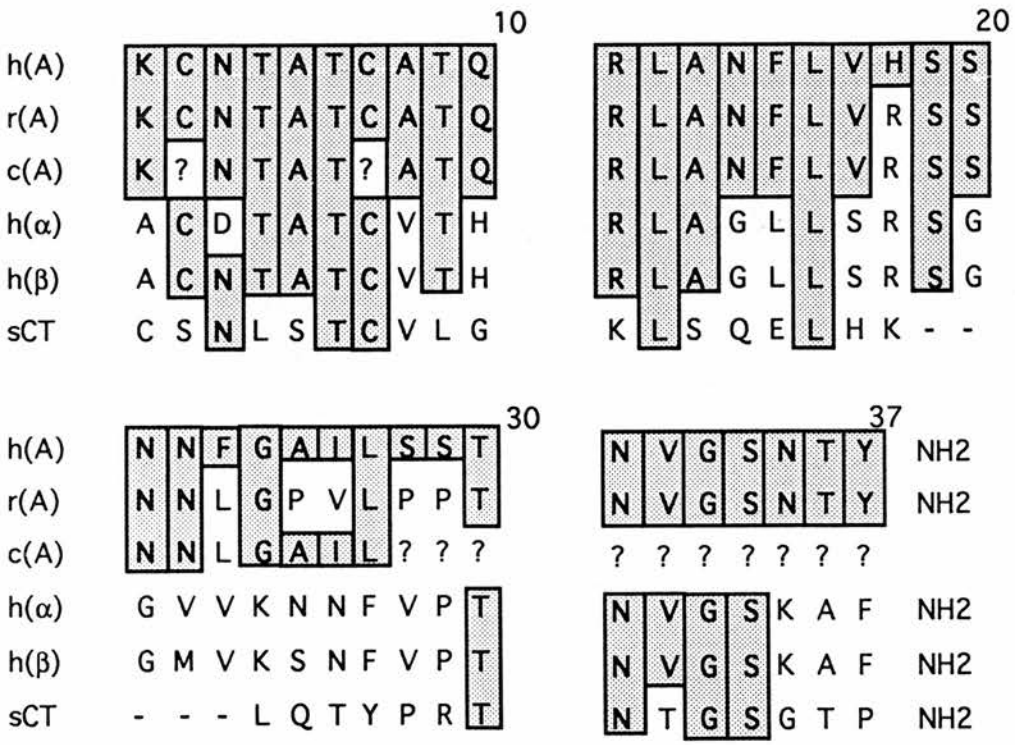
1.4 Structure of calcitonin and related peptides.

In the absence of any crystallographically determined structure most of the structural information available on the three peptides has come from spectroscopic methods. In particular, circular dichroism (CD) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy have provided the most valuable structural information.

1.4.1 Calcitonin

There are several structural features common to all calcitonins. All are 32 amino acids in length, contain a disulphide bridge between positions 1 and 7 and have a carboxyl terminal amide. A structural feature of calcitonins common to several other polypeptide hormones (Epand *et al.*, 1983) is that the primary sequence has regularly

Figure 1.8
Sequence alignment of three different Amylins with sCT, human α and human β CGRP. A gap of five amino acids has been introduced to maximise the alignment of sCT with the others. Salmon calcitonin (sCT);humanCGRP α and β, h(α) and h(β);human amylin,rat amylin and cat amylin, are h(A), r(A) and c(A) respectively. Boxed residues show sequence identity to the Amylins. Undetermined residues in the partial sequence of cat amylin are denoted with (?).



spaced hydrophobic amino acids at every third, or fourth, position in the central region of the chain. This regular spacing of hydrophobic amino acids can allow the polypeptide chain to fold into an amphipathic helix in which one face is hydrophobic and the other hydrophilic. This type of helix has been implicated in the interaction of proteins with lipids. For example, the apolipoproteins are believed to interact with phospholipid through the formation of an amphipathic helix (Segrest *et al.*, 1974). Other polypeptides that form amphipathic helices include glucagon (Jones *et al.*, 1978) and mellitin (Terwilliger *et al.*, 1982a) and it may be this property that enables these polypeptides to solubilise phospholipids. Phospholipids may be directly involved in the binding of many drugs and hormones to specific cell-surface receptor sites.

Early circular dichroism (CD) experiments proved the existence of the amphipathic helix in the presence of the acidic phospholipid dimyristoylphosphatidylglycerol (DMPG) (Epand *et al.*, 1983). Subsequent work showed that this helicity could also be induced by sodium-dodecylsulphate (SDS) (Epand *et al.*, 1985), methanol (Orlowski *et al.*, 1987) and tri-flouroethanol(Epand *et al.*, 1985). It has been assumed that this α-helix was the predicted amphipathic helix extending from Val8-Pro23. Early NMR experiments (Wüthrich, 1976) with human calcitonin in aqueous and in DMSO solutions suggested a random-coil conformation. Further NMR work on human calcitonin in DMSO and cryoprotective DMSO-water mixtures (Motta *et al.*, 1991) also did not find a helical conformation. This cryoprotective mixture has properties at low temperature similar to those of water at room temperature. In DMSO, the data indicated a highly flexible and extended conformation with both proline residues, Pro23 and Pro32,

exhibiting cis-trans isomerism. The structure was more ordered in the cryoprotective mixture and the data indicated an anti-parallel ß-sheet between Phe16 and Thr21. Only Pro23 exhibited cis-trans isomerism with 25% of the population containing the cis peptide bond at this residue. Similar NMR work has been carried out with salmon calcitonin (Motta, 1989). Data from cryoprotective mixtures of calcitonin in DMSO-water indicated an extended conformation with a short double stranded ß-sheet from Leu12-Lys18. This ß-sheet had an amphipathic distribution of residues which was postulated as the "seed" of the amphipathic α-helix responsible for the interaction with lipids, similar to the folding mechanism of signal peptide-lipid interactions (Briggs *et al.*, 1986). Once again there was cis-trans isomerism at Pro23. The proposed α-helix was finally observed in an NMR study of salmon calcitonin in 90% methanol-10% water (Meadows *et al.*, 1991). This helix extended from Val8-Tyr22 and was broken by Pro23 which exhibited cis-trans isomerism. However this was not a classic helix and an amphipathic distribution of residues in the helix was not observed (Meadows, personal communication). Additionally, the C-terminal segment was found to exhibit conformational heterogeneity. Another NMR study (Meyer, J.-P. *et al.*, 1991) performed in the presence of the helix inducing solvent trifluoroethanol identified the existence of an α-helix. Circular dichroism measurements showed that the helix was also present in aqueous solutions containing as little as 15% of TFE. Solutions containing any less TFE than this resulted in a loss of structure. This helix started at Leu4 of the N-terminal ring and had the hydrophobic face predicted by helical wheel projections (Meyer, J.-P., personal communication) as shown in plates 1.1 and 1.2. As shown in plate 1.2 this helix had an amphipathic distribution of amino

Plate 1.1
Ribbon diagram of sCT from NMR structure of Meyer et al., (1991). The helix starts at Leu 4 and terminates at Tyr 22. Proline 23 is in the *trans* conformation in this structure. Diagram produced on an ESV workstation using SYBYL.

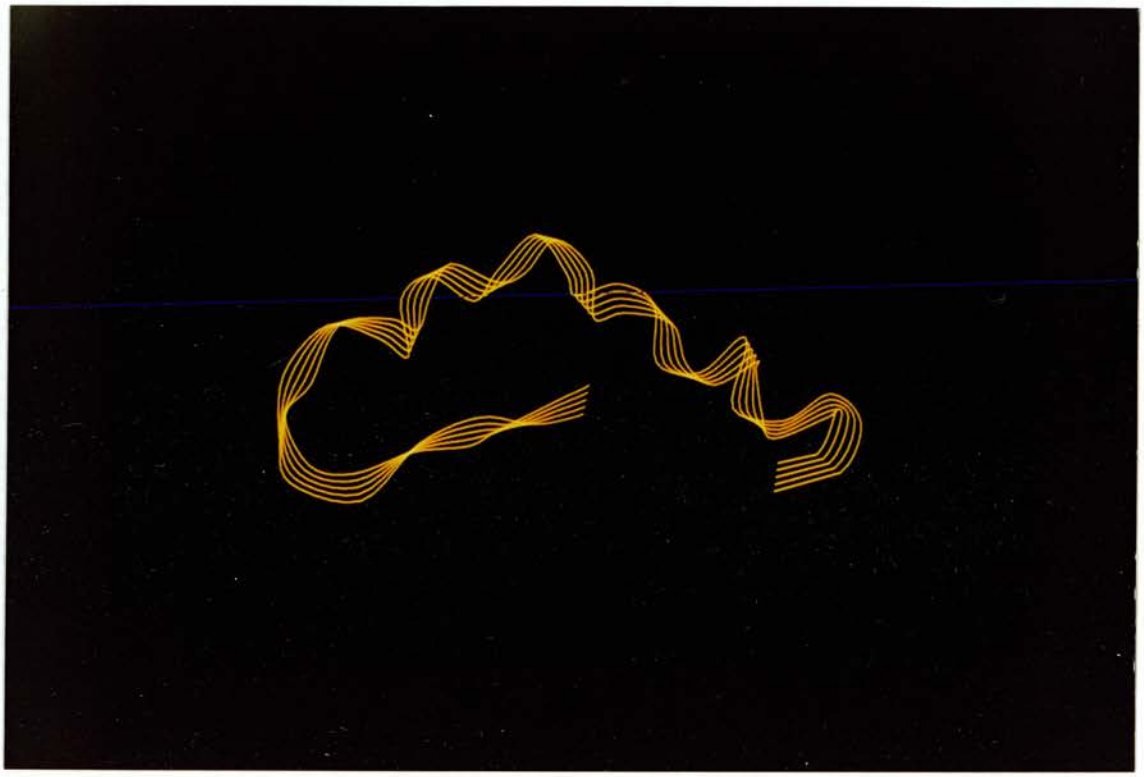
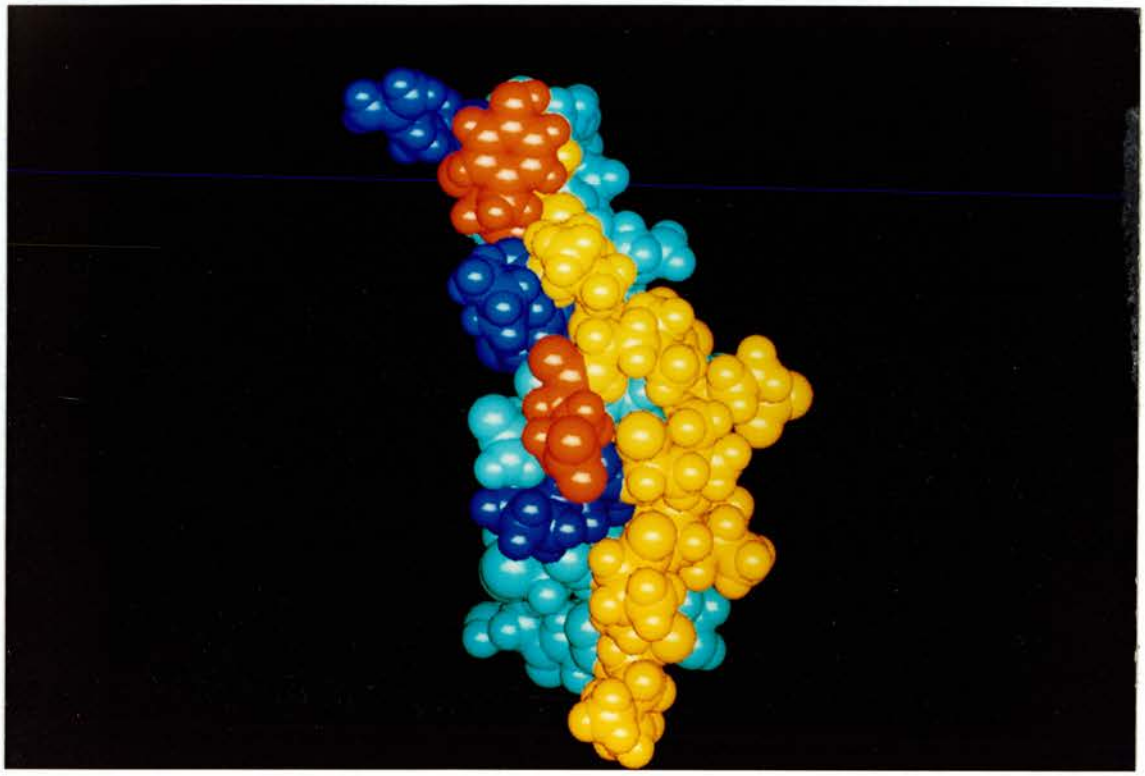


Plate 1.2
Space filling model of NMR structure from Meyer et al., (1991). In this orientation the helix axis is vertical. Hydrophobic residues are shown in orange, basic in royal blue, and acidic in red. The amphipathic nature of the helix is clearly illustrated. Proline 23 is in the *trans* conformation. Diagram produced on an ESV workstation using SYBYL.



acid sidechains. The C-terminal residues were less well assigned but a loop in the last 10 residues was identified which resulted in the interaction of the C-terminus with the helix. While it was clear that the residues in the C-terminal region were more mobile, they did possess a degree of structure that interacted with the helix. A major finding of this work was that the addition of water caused cis-trans isomerism at Pro23.

It is clear from all this NMR work that while secondary structural elements can exist, the extent and form of these are highly dependent on the conditions of the experiment. Both α-helix and ß-sheet are observed but have widely differing structures. All contain areas, primarily around the C-terminus, that are more dynamic and adopt multiple conformations.

1.4.2 Calcitonin gene-related peptide.

When the high sequence homology of CGRP with calcitonin was determined (Jonas *et al.*, 1985), subsequent work demonstrated that CGRP and sCT-I cross react with each other's receptors (Goltzmann & Mitchell, 1985). From the amino acid sequence it was suggested that CGRP, like calcitonin, had the potential to form an amphipathic α-helix between residues 8 and 25 (Lynch & Kaiser, 1988). However in aqueous solution, circular dichroism (CD) experiments have suggested that CGRP is largely unstructured with only 20% helix. This solution structure was also shown to be unaffected by pH in the range 3.5-9.0 (Hubbard *et al.*, 1991). Increased helical content has been shown to occur with 1:1 (v/v) TFE/water. Early CD work (Manning, 1989) estimated 60% helix in this TFE/water mixture whereas NMR work (Breeze *et al.*, 1991) suggested only 27%. Subsequent CD experiments (Hubbard

et al., 1991) confirmed a high helical content (45-70%) in this TFE/water mixture, and indicated a concentration dependent decrease in helicity as a possible explanation for the differences between the NMR and the earlier CD results. The NMR work (Breeze *et al.*, 1991) confirmed that an α-helix existed, which stretched from residue 8 to 18. It was also confirmed that the helix possessed an amphipathic distribution of hydrophobic and hydrophilic amino acids. The rest of the molecule was essentially disordered although there were suggestions of a turn-type conformation between residues 19 and 21. This work also indicated that the α and ß forms of CGRP were structurally very similar.

1.4.3 Amylin

As well as possessing sequence similarity with CGRP, amylin also has a structural similarity. CD experiments (Hubbard *et al.*, 1991) have suggested that amylin, like CGRP, possesses a small α-helical content (13%) in aqueous solution but this can be increased by addition of 50% v/v TFE (20%) or 0.4% SDS (62%). However unlike CGRP, amylin helix content was shown to be temperature sensitive, decreasing with increasing temperature.

X-ray diffraction studies on amyloid isolated from humans with primary and secondary systemic amyloidoses has revealed that the protein adopts a cross-ß pleated sheet conformation (Eanes & Glenner, 1968; Bonar *et al.*, 1969). It is believed that this conformation is the structural basis for both the staining properties and the fibrillar structure of the amyloids.

EX BIBL. UNIV. EDINBURGEN.

1.5 Structure-activity relationships of calcitonin and related peptides.

By measuring the biological activity of a series of related molecules it is possible to correlate increased or decreased activity with some structural feature of the molecule itself. This is the concept of the structure-activity relationship. This provides information on the nature of the interaction between a molecule and it's receptor.

1.5.1 Calcitonin

Although CT's from many species have been sequenced and characterised biologically, few correlations have been made between structure and biological activity. It has been shown that ultimobranchial calcitonins were approximately 30-fold more potent than calcitonins of thyroidal origin both *in vitro* and *in vivo* (Galente *et al.*, 1971). Analogue studies on sCT (Orlowski *et al.*, 1987) have suggested that the N-terminal disulphide-bridged ring structure is not required for biological activity. However, similar work on hCT (Rittel *et al.*, 1976), showed that hypocalcaemic activity was drastically reduced by opening of this disulphide ring. Replacement of the disulphide bridge by a carba-type analogue in the amino-suberic acid derivative of eel [ASU 1,7] calcitonin resulted in full retention of biological activity (Morikawa *et al.*, 1976). These results suggest there is not a direct relationship between the ring structure and biological activity. However it may be that the role of the N-terminal ring is to promote the formation of a helical structure in other regions of the peptide when bound to lipids. This is supported by the finding that sCT-(12-32)-peptide amide, which lacks the N-terminal ring structure, attains little helical structure even in the presence of lipid or detergent

(Epand *et al.*, 1986).

The biological activity of sCT is drastically reduced by a number of modifications at the C-terminus. The C-terminal prolinamide is absolutely required for activity. This residue is completely conserved in all the species studied so far. The position of the C-terminal amide is important as was shown in two hCT analogues, hCT-(1-31)-NH_2 and [Pro29]-hCT-(1-29)-NH_2, possessing no biological activity. Any shortening at this C-terminal end results in a loss or depletion of biological activity (Rittel *et al*, 1976). However, C-terminal fragments of eel calcitonin (eCT) have been shown to possess receptor binding activity (Yamamoto *et al.*, 1981). It has been proposed that receptor binding is a function the C-terminus which does not lead to adenylate cyclase activity. Activation of adenylate cyclase has been attributed to the N-terminus between residues 1 to 18 (Yamamoto *et al.*, 1981).

The majority of the structure activity work on calcitonins has focused on the central region (position 8-23) which has the potential to form an amphipathic helix. Attempts have been made to correlate this amphipathic helix with biological activity. However, neither the hydrophobic moment of the helix (Epand *et al.*, 1986) nor optimisation of the amphipathic helix (Moe *et al.*, 1985) result in increased biological activity. The structural requirements of this helix are not very specific and active analogues of calcitonins have been prepared with a number of substitutions in the region expected to fold into an amphipathic helix. The production of sCT analogues with increased biological activity and reduced helical content (Epand *et al.*, 1986) has reduced the significance of the amphipathic helix. Also, some analogues have been shown to retain biological activity without the

ability to form an amphipathic helix (Epand *et al.*, 1988). Thus, the structure-activity results are far from clear and in some cases appear contradictory. A knowledge of the three dimensional structure of calcitonin, or a calcitonin analogue, will greatly help the interpretation of these results.

1.5.2 Calcitonin gene-related peptide.

Existing information concerning the relationship between the structure and activity for CGRP largely hinges upon the effects of residue deletions/substitutions and of chemical modifications on biological activity. In a comprehensive study (Zaidi *et al.*, 1990) involving three different biological assays, it was shown that, as with calcitonin, the intact peptide was required for full biological activity. Even the molecule hCGRP-(Ala1-Lys35), which had lost two residues at the C terminus, resulted in a marked decrease in activity. Tryptic and chymotryptic fragments were also without activity. However, unlike calcitonin (Orlowski *et al.*, 1987), the destruction of the disulphide bridge between Cys2 and Cys7 abolished biological activity suggesting an important role for the N-terminus in the interaction of CGRP with its receptor. However, the lack of activity shown by N-terminal fragments (Ala1-Arg18) of CGRP rather suggested that the rest of the peptide was probably necessary for the peptide to assume the correct conformation for interaction with it's receptor. Also, the fact that minor modifications such as the acetylation of the lysine residues Lys24 and Lys35 or substitutions at the C-terminus, which occur naturally in ß-CGRP, do not cause significant changes in biological activity suggest an important role for the C-terminus. This was further supported by the finding that hCGRP-(Val8-Phe37) (Chiba *et*

al., 1989) possessed receptor antagonist activity. However, without the N-terminal disulphide bridge and ring structure, this C-terminal fragment was unable to induce the subsequent intracellular signal transduction event. Like calcitonin, the role of the proposed amphipathic α-helix has been investigated (Lynch & Kaiser, 1988) in CGRP. Two analogues of CGRP with differing lengths of idealised amphipathic α-helix were investigated. The first analogue, which possessed approximately five turns of helix between residue 8 and 25, had no receptor antagonist activity. The second analogue, which possessed approximately three turns of an helix between residues 8 and 18, possessed agonist activity. It was concluded that residue 18 represented the termination point for the amphipathic α-helix. This was also the finding of the NMR study carried out on CGRP (Breeze *et al.*, 1991). It is therefore highly likely that this amphipathic α-helix plays an important role in the interaction of CGRP with it's receptor.

1.5.3 Amylin

Since amylin's role is still not fully understood little work has been carried out on structure activity relationships. Molecular modelling of amylin (Saldanha & Mahadevan, 1991) using a combination of knowledge sources, has suggested an α-helix/ß-sheet motif similar to that found in the insulin B-chain. This work has suggested a possible explanation for the aggregation properties of amylin. Hydrophobic patches which are at residues 13, 15, 16, 17 in the α-helix and residues 23, 25, 26, 27 in the ß-strand are thought to be important in fibrillogenesis and polymerisation of amylin. The hypothesis that hydrophobic patches are important in the

fibrillogenesis is that in the rat (Leffert *et al.*, 1989) amylin does not form amyloid. This has been shown to be related to a changed sequence motif in this species (Betsholtz *et al.*, 1989). In the modelled rat amylin the C-terminal ß-strand sequence suggested a polyproline-like helix sequence and was modelled as such rather than a ß-strand.

1.6 Overall Similarities between calcitonin and related peptides.

It is clear that these polypeptides are each capable of carrying out a wide variety of biological activities. However, it is also clear that they have certain biological activities in common. For example, CT and CGRP can cross react at each other's receptors ultimately affecting calcium levels. Both CGRP and amylin can inhibit *in vitro* glycogen synthesis. These functional similarities are presumably a result of some structural similarity. This biochemical evidence is supported by the sequence similarities shown in figures 1.7 and 1.8. The possibility that all three peptides are related structurally has never before been proposed. Structure prediction and structural work on all three peptides has indicated an important role for an amphipathic α-helix/turn structure and this may be a common motif for all three of these peptides. It is highly likely that this structure is responsible for the interaction of both calcitonin and CGRP with their respective receptors. This idea is further supported by the observation that calcitonin and CGRP can cross react at each others receptors while possessing amphipathic helices of approximately the same length. Thus a structural understanding of one of these, calcitonin, will aid in understanding the structure/activity similarities of all three. In an attempt to use x-ray crystallography

to provide this structural information, Chapter 4 outlines the first step in this process - the attempted crystallisation of sCT and two derivatives. Having determined the crystal structure of calcitonin it may subsequently be possible to rationally design more active analogues or simpler analogues which possess the necessary biological activities.

Chapter 2 Small molecule crystallography

Chapter 2

2.1 Introduction.

The purpose of this Chapter is to give a brief introduction to the theory of the methods used in the structural determination of a small biologically active molecule. Most of this information forms the basis for similar work in protein crystallography. For a fuller description the reader is directed to Stout & Jensen (1968) and Woolfson (1970). The structure of this molecule is compared to the structures of previously determined structures of similar biological activity. It is hoped that important structural features of these molecules can be identified.

2.1.1 Diffraction of X-rays.

Diffraction is the phenomenon which occurs whenever a wave motion interacts with an obstacle. The first diffraction of X-rays by a crystal in 1912 led to the final proof of the wave nature of X-rays. However it was not until Bragg noted the similarity of diffraction from a crystal to ordinary reflection and derived a simple equation treating diffraction as "reflection" from planes of atoms, that X-ray crystallography became tractable.

If we consider two parallel planes P_1 and P_2 with interplanar spacing d, illuminated by an x-ray beam at an angle Θ as shown in Figure 2.1 then the parallel incident rays 1 and 2 will cause electrons (assumed at O and C) to vibrate by the oscillating field of the incident beam and, as vibrating charges, will radiate in all directions. Constructive interference between rays scattered from successive planes in the crystal will only take place if the path difference between the rays is equivalent to an integral number of wavelengths

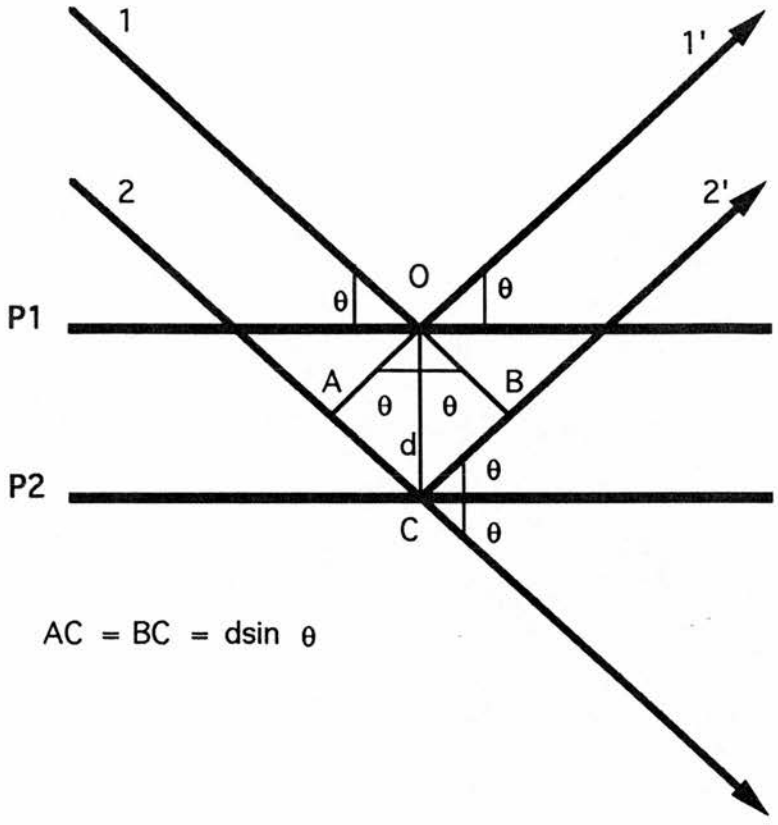


Figure 2.1 Construction showing the conditions for diffraction from a pair of parallel planes P1 and P2 with interplanar spacing d. The parallel rays 1 and 2 are incident on these planes with an angle θ.

Thus the derived equation for diffraction to occur is

$$n\lambda = 2\,d \sin\Theta \qquad (1)$$

With reference to Figure 2.1 drawing perpendiculars from O to A and B, respectively, it is clear that $<AOC = <BOC = \Theta$. Therefore AC = BC and waves in 2' will be in phase with 1' if the path difference AC + BC (= 2AC) is an integral number of wavelengths, λ. It can also be shown that what holds for points O and C can be shown to hold for any points in planes 1 and 2. Therefore any ray parallel to 1' but from another point in plane P_1, will differ in phase by $n\lambda$ not only from ray 2' emergent from C but also from any other parallel ray from any other point on plane P_2.

In the derivation of Bragg's law electron density is described as a plane. In crystal structures, the electron density is distributed throughout what is called the unit cell. Bragg's law is still valid in the crystal, since waves scattered from electron density not lying in the plane P may be added to give a resultant as if reflected from the plane. Variation in these resultants accounts for the differing intensities of reflection observed for various planes. Diffraction maxima from the crystal are sharp and occur only at clearly defined values of 2Θ because of the fact that hundreds or thousands of planes make up each of the mosaic blocks.

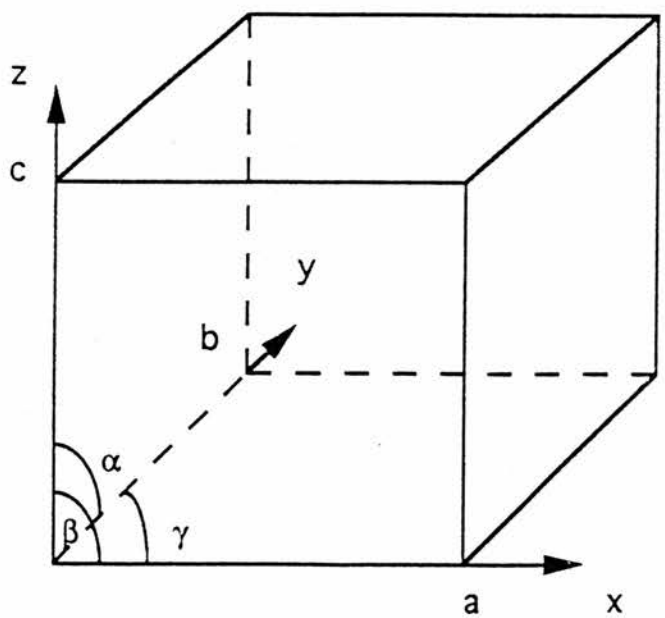
2.2 Crystal Systematics.

2.2.1 The Unit Cell.

Since the seventeenth century, it has been realised that crystals contain a high degree of internal order. A detailed study of crystals led to the realisation that regularity of form of different crystals must result from an ordered packing within the crystal (Burke, 1966).

This led von Laue to suggest that crystals could be treated as a three-dimensional diffraction grating when radiation of a comparable wavelength to the interatomic distances within the crystal was used. Thus any crystal can be regarded as being built from blocks of a unit structure regularly repeated in three dimensions. The unit structure may be one or more atoms, a molecule, or a complex assembly of molecules. This unit structure is termed the Unit Cell and is defined as the basic parallelipiped shaped block from which the whole volume of the crystal may be built by regular assembly of these blocks. Unit Cells are chosen, by convenience, to have the shortest edges possible and at the same time be in accord with the highest possible symmetry of the crystal. The edges are defined by three vectors, which are conventionally given the symbols a, b and c and the angle between these defined as α, ß and ɣ as shown in Figure 2.2.

Figure 2.2 The Unit cell showing how the cell edges and angles are defined.



The axial lengths are a, b and c. Any point in the unit cell can be specified with respect to a given origin by parameters x, y and z measured parallel to the axes and expressed as fractions of the cell edges.

2.2.2 The Space Lattice.

If each unit cell is represented by a point at some defined position, then the array of cells can be defined by a lattice of points. Each point is in exactly the same environment and orientation as every other point. This is the space lattice. The concept of the lattice is important when diffraction by a crystal is considered. Only 14 distinct kinds of lattices - the Bravais lattices - may be geometrically constructed which satisfy the condition that the environment and orientation of each lattice point is the same. These lattices have one of seven different unit cell shapes with different symmetry properties, corresponding to the seven crystal systems. The arrangement of structures on these lattices must be consistent with one of the 230 different combinations of symmetry elements that are possible for arranging objects in a regularly repeating manner in three dimensions. These are the 230 space groups. All 230 space groups and the systematically absent reflections that they imply are listed in International Tables, Vol.A (1983). Therefore before any crystal structure can be determined it is necessary to determine the underlying symmetry of the crystal - its space group. The symmetry of the crystal will be manifest in the symmetry of the diffraction pattern leaving only the asymmetric portion of the unit cell, the asymmetric unit, to be determined. Although representatives of most, but not all, of the 230 space groups have been found, about

60% of the organic compounds studied crystallise in one of six space groups.

In the explanation of Bragg's Law, diffraction was described as reflection from planes. Sets of parallel planes can be constructed for the lattice, so that for any given set of planes every lattice point lies on some member of the set. These sets of planes are identified by three indices, one corresponding to each axis. The intercept of the plane with the cell axes are expressed as fractions of the unit cell edge, their reciprocals are just the desired indices. Thus the set of planes defined by h,k,l intercept the unit cell axes at a/h, b/k, c/l where h,k,l are three integers. The h, k, l are also the orders of diffraction with respect to each axis.

2.2.3 The Reciprocal Lattice.

The reciprocal lattice is a necessary extension of the space lattice from which it is derived. Considering Bragg's law in the form

$$\sin \Theta = \frac{n\lambda}{2d} \qquad (2)$$

sin Θ is inversely proportional to d, the interplanar spacing. To make this relationship a direct one and allow easy interpretation of diffraction patterns a reciprocal lattice is constructed, based on 1/d which varies directly as sin Θ. The reciprocal lattice can be defined by considering normals to all possible direct lattice planes (h, k, l) to radiate from some lattice point taken as the origin. Each normal is terminated at a point a distance $1/d_{hkl}$ from this origin, where d_{hkl} is the perpendicular distance between the set of planes (hkl). The set of points so generated constitutes the reciprocal lattice. As before, sets of parallel planes can be constructed so that for any given set,

every reciprocal lattice point lies on some member of it. The planes of these reciprocal lattice points are now directly related to the space lattice axes. Reciprocal cell lengths are designated by a^*,b^*,c^* and the angles α^*,β^*, γ^*. For lattices that are either orthorhombic, tetragonal or cubic the algebraic expressions relating a parameter of the reciprocal unit cell to the real unit cell are simple, for example $a^* = 1/a$. For the other lattice types the algebraic relationships are more complex. The most important property of the reciprocal lattice is that it allows easier understanding of Bragg's law in the diffraction experiment. For a more comprehensive treatment of the reciprocal lattice the reader is directed to Woolfson (1970).

2.2.4 Crystallisation.

These concepts of planes and lattices are the imaginary constructs that are useful for describing diffraction from crystals. However, obtaining a suitable crystal for study and the collection of data represent the practical considerations. Protein crystals and the crystallisation of proteins will be discussed separately in Chapter 4. The theory of crystallisation is applicable to all molecules, only the practical aspects differ between protein and small molecule crystallisation. Briefly the crystallisation of a molecule represents the dynamic equilibrium between the molecule in the fluid phase and in the solid phase. The crystal is highly ordered with the molecules in regular positions. The loss of the entropic energy of the molecules is compensated for by the various types of bonding energies formed by the ordering process. The growth of most organic and inorganic crystals is achieved by producing a supersaturated solution of the compound in a suitable solvent. A supersaturated solution can be obtained in a

number of ways namely by slow evaporation, slow cooling and diffusion. The crystallisation of small molecules is relatively simple compared to protein crystallisation. For small molecules, crystallisation generally occurs very rapidly resulting in microcrystalline material which is not suitable for analysis. Thus the main problem for small molecule crystallisation is controlling the process to produce large single crystals. Slow evaporation can be used when the solvent of the saturated solution is slowly allowed to evaporate. This method requires the use of volatile solvents, evaporation rates can be controlled and best results occur when the rate is slow. Slow cooling requires that a solution of the compound is made at an elevated temperature which on cooling becomes supersaturated. The compound must be stable at the elevated temperature if this approach is to be used. Here also the rate of cooling can be controlled and best results are normally when the rate of cooling is slow. If difficulty is experienced in growing suitable crystals from single solvents, mixtures of two or more solvents can be used. The properties of the solvent, such as density, dielectric constant, viscosity, solvation and solubility can have a pronounced effect on crystal growth. These can be varied in a controlled way over wide ranges and adjusted to fit the particular circumstances.

2.2.5 Choice of Crystal.

If a crystal is to be of any use for collecting x-ray diffraction data it has to fulfil two main requirements:(1) it must possess uniform internal structure and (2) it must be of proper size and shape. Crystals with any physical defects such as being twinned or cracked are of no use. It is not necessary for the crystal to possess

uniform or well defined external faces. The ability of a crystal to polarise light under a polarising microscope can be used to screen suitable crystals. In order that all parts of the crystal are exposed to the same intensity of radiation the crystal should not be too large. If the crystal exceeds about 0.5mm x 0.5 mm it may need to be cut or shaped. However, some crystals are too soft, fragile, or sensitive for cutting that they must be used as they are grown. Since the intensities of diffracted rays are proportional to the amount of material that is scattering them, crystals that are too small are generally not suitable. The absorption of x-rays is an important factor concerning the shape of the crystal. If the crystal is particularly thick in one dimension then the opportunity for absorption of diffracted rays will be increased relative to other reflections. As a result a systematic error will be introduced into the observed intensities. It is possible to make a calculated correction for this effect, but the corrections are not very accurate and it is therefore better to try and avoid the problem by proper choice of the crystal.

2.3 Intensity data collection.

There are two general methods for measuring intensities of diffracted rays. Either the ray can be detected by a quantum counting device measuring the photons directly (diffractometers), or else the degree of blackening of photographic emulsion may be measured and taken as directly proportional to the intensity of the diffracted ray. Automated diffractometers are the more recent development and have been used in the data collection for the work described in this thesis. Photographic methods will not be discussed. However, the

corrections and the production of integrated intensities are essentially the same for both types of data collection.

Integrated intensities are a measure of the total number of photons of the characteristic wavelength diffracted when the Bragg condition is satisfied. However the actual measurements made are only close to these values. Certain corrections have to be made to produce these integrated intensities. Difficulties of intensity measurement arise from three sources:

1. Scattering of radiation by processes other than Bragg reflection.
2. The spectral impurity of the beam.
3. Diffraction spot size and shape variation.

For true intensities the background radiation must be removed. There are two sources of background radiation, the first from diffuse scattering of x-rays by the crystal itself and by air, the second arises from spectral impurities and can be removed by selective filtering. Two reflections of the same integrated intensity can have widely differing spot shapes and sizes and also different intensities at some defined point within each spot. The reasons for this are partly due to apparent changes in the illuminated area of the crystal surface as the crystal rotates and also to geometric factors. There is a general tendency for reflections to be larger and more diffuse at high values of sin Θ than at low values.

The most common instrument for automatic intensity data collection is the four-circle diffractometer as shown in Figure 2.3. The four circles of rotation allow the crystal to be oriented in such a way as to bring the normal of any desired plane (i.e. the scattering vector) into the equatorial plane where it can be detected by the counter

which is moved to the Bragg position by the 4th circle. The entire instrument is controlled by a computer and the angular settings of the crystal and the detector can be calculated from the unit cell dimensions and the orientation of the crystal on the instrument. Thus the whole data collection can be done automatically.

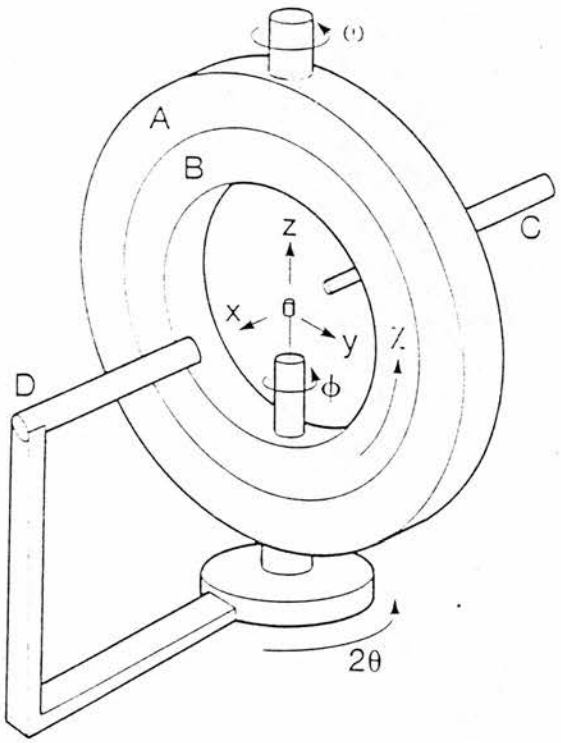


Figure 2.3 The Four circle diffractometer. The crystal is randomly mounted on the spindle which can be rotated around the ϕ-axis. This spindle assembly can be rotated about a perpendicular axis, the χ axis, by rotation of ring B inside the stationary ring A. This χ-ring together with the spindle assembly and the crystal can be rotated about a vertical axis, the ω-axis. The detector D is mounted on an arm which enables it to be rotated about a vertical axis, designated the 2θ-axis, coincident with the ω-axis. C represents the collimator.

Standard reflections are measured regularly so that crystal slippage and radiation damage (which would result in weaker diffracted intensities) can be observed.

2.4 Data reduction.

The result of the collection of X-ray diffraction data is a relative intensity, I, for each reflection with indices, h, k, l, together with the scattering angle, 2Θ, for that reflection and some estimate in the error in the measurement of the reflection (standard deviation). Intensities are not the most suitable form in which the measured data can be expressed. The required quantity, which can be derived from intensity, is the structure factor modulus $|F_{hkl}|$. This is related to the observed intensity by

$$|F| \propto I^{1/2} \qquad (3)$$

The structure factor has both magnitude (amplitude) and phase (relative to the origin of the unit cell). Once the approximate positions of the atoms in the cell are known these structure factors can be theoretically calculated. Structure factors are also used in the calculation of electron density maps from which the position of atoms may be determined. Thus it is essential to convert these intensities into observed structure amplitudes $|F_o|$ for use in subsequent calculations.

2.4.1 Lorentz and polarisation corrections.

As described previously, certain corrections, primarily geometric, have necessarily to be applied to $|F_o|$. The first correction is the Lorentz factor L_{hkl}, which corrects for the different geometries of intersection of a lattice point with the Ewald sphere when it satisfies the Bragg condition. This is described for a four circle diffractometer, with the geometry as shown in figure 2.3, by the factor

$$L = 1/\sin 2\Theta \qquad (4)$$

however other geometries have different expressions for the Lorentz factor and can be found in Vol.IV of The International Tables for X-ray Crystallography (1974). The second correction is a polarisation correction (p) which corrects for the partial polarisation of the diffracted beam. This is also an angle dependent term described by

$$p = \frac{1 + \cos^2 2\Theta}{2} \qquad (5)$$

and is independent of the method used for data collection. Thus the result of these data reduction corrections are $|F_{relative}|$'s ($|F_{rel}|$) defined by

$$|F_{rel}| = (I_{hkl}/Lp)^{1/2} \qquad (6)$$

The output data after data reduction therefore contains hkl, $|F_{rel}|$, $(\sin\Theta)/\lambda$ and some measure of the probable error in the $|F_{rel}|$.

2.4.2 Atomic scattering factors.

The scattering power of a given atom for a given reflection is known as its scattering factor (f_o) and is expressed in terms of the scattering power of an equivalent number of electrons located at the position of the atomic nucleus. The scattering factor is always equal to the total number of electrons in the atom at $\sin\Theta/\lambda = 0$ and decreases as $\sin\Theta/\lambda$ increases. Scattering factor curves have been calculated for a large number of atomic species using theoretical electron distributions. The results are available in the International Tables for X-ray Crystallography Vol.IV (1974).

2.4.3 Temperature factors and absolute scaling.

In the solution of the crystal structure the atoms are assumed to exist at discrete positions within the unit cell. This does not take

account of the vibrational motion of atoms at non absolute zero temperatures. Thermal energy causes the atoms to oscillate about their equilibrium positions. The effect of this is to reduce the effective scattering power of an atom. The change in scattering power has been theoretically shown to be given by

$$e^{-B[(\sin^2\Theta)/\lambda^2]} \tag{7}$$

where B is related to the mean-square amplitude (u^2) of atomic vibration by

$$B = 8\pi^2 u^2 \tag{8}$$

Thus the true scattering factor for an atom becomes

$$f = f_o e^{-B[(\sin^2\Theta)/\lambda^2]} \tag{9}$$

This assumes the amplitudes of vibration are radially symmetric. However the most realistic model for atoms in a crystal is to treat the vibrations as ellipsoidal rather than as a sphere. There are six parameters for the ellipse which define the lengths of the axes and their orientation relative to the crystal axes.

An estimate of the overall temperature factor can be determined from the data by means of the Wilson plot. This plot also gives scaling information relating the average observed intensity to the theoretical value predicted knowing the number and type of atoms in the unit cell. This absolute value of intensity is calculated in small shells of ($\sin^2\Theta/\lambda^2$) and compared to the observed. The thermal motion is calculated by

$$I_{abs} = \Sigma f_o^{\,2} e^{-2B(\sin 2\Theta)/\lambda^2} \tag{10}$$

where B remains to be determined. The assumption which is necessary is that the average intensity of diffraction at a certain 2Θ value depends only on the contents of the unit cell and not their positions. Thus a straight line plot of $\ln(I_{rel}/\Sigma f^2)$ against

$\sin 2\Theta/\lambda^2$ allows B to be derived from the slope and the scale to be derived from the intercept when $\sin 2\Theta/\lambda^2 = 0$. The resulting scale factor can be applied to the data to put the values of $|F|$ on an approximately absolute scale for all reflections measured.

2.5 The Fourier syntheses and diffraction.

2.5.1. The structure factor.

As previously mentioned, the calculation of the structure factor is an important step in the determination of any structure. The Structure Factor, F_{hkl}, is the resultant of j waves scattered in the direction of the reflection hkl by the j atoms of the unit cell. It is measured relative to the scattering of a single electron at the origin and is the Fourier transform of the scattering density sampled at reciprocal point hkl. The structure factor can be represented in either of two mathematical ways which describe a wave, namely exponentially or as an ordinary complex number.

In the exponential form,

$$F_{hkl} = \Sigma_j f_j e^{2\pi i(hxj + kyj + lzj)} \qquad (11)$$

$|F|$ or $|F_{hkl}|$ - Amplitude of the scattered wave.

In the complex number form,

$$F(hkl) = A(hkl) + iB(hkl) \qquad (12)$$

$$A(hkl) = \Sigma f_j \cos 2\pi(hx_j + ky_j + lz_j)$$

$$B(hkl) = \Sigma f_j \sin 2\pi(hx_j + ky_j + lz_j)$$

Where x,y,z are expressed as fractions of the unit cell lengths a,b and c.

$i =$ $-1^{1/2}$ "imaginary number"

$f_j =$ atomic scattering factor amplitude

The phase, α, of the scattered wave can be determined by

$$\alpha = \tan^{-1}(B/A) \quad (13)$$

and

$$| F | = (A^2 + B^2)^{1/2} \quad (14)$$

2.5.2. Fourier syntheses.

Fourier's theorem states that a continuous, single-valued, periodic function can be represented by the summation of cosine and sine terms. Since crystals represent the periodic distribution of scattering matter the Fourier Series can be used. Just as the diffraction pattern is a Fourier summation of the electron density (in the calculation of structure factors), the electron density is a Fourier summation of the structure factors. The required series to describe the electron density of a point (x,y,z) in the crystal is of the form,

$$p(x,y,z) = 1/V \; \Sigma_h \Sigma_k \Sigma_l \; | F_{hkl} | \cos 2\pi(hx+ky+lz - \alpha'_{hkl}) \quad (15)$$

where V - the volume of the unit cell

p(xyz) - the electron density at a point x,y,z

α'_{hkl} - the phase angle

2.6 The Phase problem (the direct methods approach).

If | F | is known and α (for each h,k,l) can be calculated, then the electron density for all values of x, y and z can be determined. However, while | F | can be measured (from the diffraction intensities), the phase angle α is lost during the diffraction experiment. This is the essence of what is termed "The Phase Problem". Therefore values of α must be derived by some other technique. There are a number of techniques that can be used to determine the phase information. However only the Direct Methods approach will be discussed in detail

since it was used in the work for this thesis. This is a method that is still only generally applicable to small molecule structure determination. Details of the other methods of phase determination are discussed in Stout & Jensen (1969).

The Direct methods approach is so called since it derives a set of phases by consideration solely of the structure factor amplitudes. The intensity data, or the structure factors $|F(hkl)|$ are converted to normalised structure factors $|E(hkl)|$. The normalised structure factors are produced as $E^2(hkl) = F^2(hkl)/\langle F^2 \rangle$ where $\langle F^2 \rangle$ is the mean of F^2 for ranges of $\sin\Theta$. The method relies on three general conditions (i) the electron density is everywhere positive, (ii) the atoms are assumed to be spherically symmetrical and have the same shape, (iii) the structure consists of identical atoms fully resolved from each other (approximately true for organic crystals). Such conditions, apart from any knowledge of the relative positions of the atoms, are sufficient to place some limitations on the phases of the reflections. There are important probability relationships that apply between groups of structure factors. The most important of these is the $\Sigma 2$ relationship: for large values of $|E|$, say $|E_1|$, $|E_2|$ and $|E_3|$ all >1.5,

if: $h_1+h_2+h_3 = k_1+k_2+k_3 = l_1+l_2+l_3 = 0$ then: $\phi_1+\phi_2+\phi_3 \approx 0$

Direct methods are generally applicable to all small molecule structure solutions and these probabilistic equations have been included in a number computer programmes for structure determination. It has been particularly useful for centrosymmetric space groups, where the phase of each reflection is either 0 or 180°, although more recently non centrosymmetric space groups can be solved by direct methods. The result of this method is a starting structure which can

be used for the refinement process.

2.7 Refinement.

The refinement method that will be discussed here is that of Least Squares. It is a method of great power and generality and it was the method used in the refinement of both the small molecule and the protein equine PLA-2 in this thesis. It is a technique for adjusting a set of variables (called the parameters) in such a way that they give the best fit between the measured experimental observations and those calculated from some assumed model. The best parameters for the assumed model are obtained by minimising the sum of the squares of the differences between the experimental quantities and the values of the same quantities calculated with the derived parameters. Thus the function to be minimised is

$$R = \Sigma_i \, w_i(y_o - y_c)_i^{\,2} \qquad (16)$$

where y_o and y_c are the measured and calculated values for a set of n observations of weight w, with i running from 1 to m. If the calculated values are defined in terms of a set of m parameters, p_j, this minimum occurs when:

$$dR/dp_j = \Sigma_i \, -2w_i(y_o - y_c)_i dy_c/dp_j = 0 \qquad (17)$$

for all m values of j. These equations are called the Normal equations for the refinement. The method is only valid if the observations are independent of one another and that there are at least five times as many observations as parameters ($n >> m$).

2.7.1 Refinement of cell parameters.

A solution to these equations can be found if the relationship between the observations and the parameters is linear. This is the

method that is used in the refinement of the cell parameters. A set of accurately measured values of sin Θ are measured for between 20 and 30 strong reflections. These form the basis for the observational equations of the form:

$$4\sin^2\Theta/\lambda^2 = h^2a^{*2}+k^2b^{*2}+l^2c^{*2}$$
$$+2klb^*c^*\cos\alpha^*+2hla^*c^*\cos\beta^*+2hka^*b^*\cos\gamma^* \quad (18)$$

The equations are linear in terms of the six parameters a^{*2}, b^{*2}, c^{*2}, $b^*c^*\cos\alpha$, $a^*c^*\cos\beta$ and $a^*b^*\cos\gamma^*$. Unit weights are used and the resulting six linear equations in six unknowns are solved by matrix methods. The use of matrix methods allows the calculation of errors in the parameters. These are obtained from the diagonal terms of the inverse matrix which are directly proportional to the required errors. Also the off-diagonal terms provide information about the correlation between the parameters. The correlation between two parameters p_j and p_k will be $I_{jk}/(I_{jj}\ I_{kk})^{1/2}$. If the parameters are genuinely independent, this will be 0.

2.7.2. Structure refinement.

The quantity to be minimised in structure refinement is:

$$D = \Sigma_{hkl}\, w_{hkl} (|F_o| - |F_c|)^2 \quad (19)$$

where w_{hkl} is the weight of the observation and Σ_{hkl} indicates summation over all reflections. If the relationship between the observations and the parameters is not linear, as is the case here, least squares cannot be used directly to refine the structure. However it can be used in the refinement by making use of the approximate

relationship, neglecting the higher order terms in the Taylor expansion:

$$\text{for } y = f(p1,p2,p3.....)\ \delta y \approx \Sigma_j\ \delta p_j.dy/dp_j \qquad (20)$$

Thus, providing reasonably approximate values are known, there is an approximate linear relationship for shifts in parameters. Since the results of the calculation are only approximate a cyclical arrangement is needed such that the new parameters serve as the starting point for the next calculation. The cycles continue until the shifts in the parameters are less than the errors associated with the parameters themselves.

A measure of the correctness of the model structure is the fit of the observed structure factor amplitudes to those calculated from the model. This is called the conventional residual, R, which is calculated over all reflections:

$$R = \frac{\Sigma\ |\ |F_o| - |F_c|\ |}{\Sigma\ |F_o|} \qquad (21)$$

Values for R below 0.06 (6%) are regarded as reliably determined structures. However sensible stereochemistry in the model is essential as well as a low R value.

2.8 Introduction to Drug 212-842.

This compound, 4-[4-{2-(3,4-dimethoxyphenyl)ethyl}-1-piperazinyl]-1H-indole-2-carboxylic acid methyl ester, here termed 212-842, is a small biologically active molecule. The compound has blood pressure lowering

activity by acting as an antagonist to both α- and ß-adrenoreceptors. These receptors are classified into α1, α2, ß1, ß2, distinguishable on the basis of their affinities for different agonists and antagonists and have different tissue specific effects. Stimulation of α-adrenoreceptors by adrenalin, the natural agonist, causes arteriole contraction. Stimulation of ß-adrenoreceptors by adrenalin or noradrenalin causes increased rate of contraction in heart cells. In addition to this the compound has partial activity, by acting as an agonist, at 5HT receptors. Like the adrenoreceptors there are multiple populations of 5HT receptors and these have been identified as 5HT1, 5HT2 and 5HT3. Stimulation of these 5HT receptors causes vasoconstriction and smooth muscle contraction. It has also been shown (Glennon, R.A. 1987) that there is heterogeneity at the 5HT1 receptor (i.e., $5HT_{1A}$, $5HT_{1B}$, $5HT_{1C}$ and $5HT_{1D}$ sites). The compound 212-842 has specificity for $5HT_{1A}$ receptors. Thus 212-842 has a wide range of activities which are of potential clinical application.

2.8.1 Experimental.

Large rectangular crystals were obtained from a supersaturated solution of the drug 212-842 in malonic acid at 297K. A suitable single crystal of dimensions 0.7 x 0.65 x 0.4 mm was selected for data collection. The crystal was mounted with its longest edge along the axis of the spindle. A unit cell was determined and refined using 25 strong reflections. The space group was monoclinic, $P2_1/a$, with unit cell dimensions a = 15.9531(6), b = 9.9683(7), c = 16.9717(9) Å, ß = 91.4517(5)° (the numbers in the brackets refer to the estimated standard deviations in the last place quoted). Thus the unit cell volume is 2698.9 $Å^3$ and contains 4 molecules giving the crystal a

calculated density (D_x) of 1.299 gcm^{-3}.

The data set was collected on a Enraf-Nonius CAD-4 using Nickel filtered Cu Kα radiation (λ = 1.4518 Å). The range of indices collected was $-17 < h < 17$, $0 < k < 11$, $-1 < l < 19$, Θ_{max} was 60°. The standard reflections chosen to monitor crystal decay and slippage were 0 4 0 and 10 0 0. No significant variation in intensity was observed during data collection and no absorption corrections were applied to the data. During data collection 4429 reflections were measured. The data for the refinement were sorted and merged to produce 3993 unique reflections, once all the systematically absent reflections were removed. The E-statistics as a function of sin Θ were close to unity indicating a centrosymmetric structure. The number of reflections used in the further refinement were reduced to 3663 after 330 reflections with $I < 2\sigma$ were suppressed.

The location of non hydrogen atoms was achieved using direct methods with SHELXS86 (Sheldrick,1986). Two molecules, namely the drug and the malonic acid in which the drug was crystallised, were located. The coordinates of these were used in the subsequent refinement process which was carried out using SHELXS-76. The R factor at this stage was 0.212 and after 4 cycles of least squares including all the non hydrogen atoms with isotropic temperature factors the R factor dropped to 0.146. At this stage hydrogens were added to their calculated positions and allowed to ride on the atoms to which they were attached, with fixed isotropic temperature factors and site occupancies. No hydrogens were placed on the carboxyl groups of the malonic acid. Finally all the non hydrogen atoms were refined with anisotropic temperature factors. The R factor dropped to 0.053. At this stage the difference map contained two additional peaks, each of

a size equivalent to an electron, that were not accounted for by the structure. These were in approximately correct positions to be assumed to be hydrogen atoms. The first of these was within 0.98 Å from the piperazine nitrogen N(6). The other was within 1.02 Å from the carboxyl O(37) of the malonic acid. These were included as hydrogens with fixed occupancies and isotropic temperature factors before a further 4 cycles of refinement. However after this refinement the solvent hydrogen H(37 O) - O(37) distance was 0.79 Å. This bond was therefore constrained to be a distance of 1.0 Å in subsequent refinements. After a further 4 cycles of refinement the R factor was 0.0493. It was at this stage that unit weights were dropped in favour of those based on σ(F). The scheme in SHELX-76 uses

$$\text{weight} = K / (\sigma^2 (F) + gF^2) \qquad (22)$$

(K and g are constants which are redetermined after each structure factor calculation) the value chosen initially for g was -0.0001 but both K and the absolute value of g are refined when g is given as a negative quantity. The weighting scheme was refined for six cycles. The atomic parameters were not refined. The refinement proved to be no better than unit weights. The initial value of g was reduced to -0.000001 and the values refined for another six cycles as before. While there was a very slight improvement R = 0.0492 this weighting scheme was finally dropped in favour of unit weights. Ten reflections with $\delta/\sigma > 6.0$ were suppressed and the coordinates refined for 2 cycles. The R factor was now 0.0488. While the structure was now solved the difference map contained a trough twice the size of the largest peak suggesting more than random noise. A further 3 reflections with $\delta/\sigma > 6.0$ were suppressed and 4 cycles of refinement did not remove the trough. It was necessary to determine the

explanation for this trough in the difference map. Thus a difference map was produced from SHELX-76 which was converted to a format suitable for FRODO. Both the map and the structure were viewed on an Evans & Sutherland PS300. The map was contoured so that this negative density could be viewed. From this it was easily seen that the holes in the map were associated with the hydrogen atoms of three methyl groups. Thus the methyl groups of C (1), C (27) and C (30) were subsequently refined as rigid groups for a further 4 cycles. The maximum shift (δ/σ) in the final run was 0.07. The final R factor was 0.0479. During refinement there were 343 reflections with $I < 2\sigma$, these were treated as unobserved. The total number of parameters refined was 355, the number of observations used was 3663. The final difference map had maximum and minimum values of + 0.19 and - 0.298 $eÅ^{-3}$ respectively and the maximum $\sin\Theta/\lambda$ was 0.623 $Å^{-1}$.

The program CALC (Gould and Taylor, 1983) was used to provide the molecular geometry data presented in table 2.2a-c and also to provide interplanar angles and centre of gravity values in table 2.5. Diagrams were prepared using PLUTO (Motherwell, 1972) and an interactive version of ORTEP (Johnson, 1965:Mallinson & Muir, 1985). Atomic scattering factors were from International Tables for X-ray Crystallography Vol.IV (1974). The final atomic parameters are given in table 2.1a-b. A connectivity search was carried out on the Cambridge Structural Database (Allen et. al., 1979) for related structures. The fragment chosen for the search was a benzyl ring bonded to a piperazine ring through the nitrogen atom. This search retrieved 33 compounds from the database of which 2 compounds were selected which possessed similar activity to 212-842. These two compounds together with 212-842 are shown in figure 2.6. The first

compound QUEST88 REFCODE = COKTAX possess ß blocking activity, the second QUEST88 REFCODE = CAGXIR has α-adrenoreceptor antagonist activity and is a tranquiliser. Molecular superposition and graphics visualisation was carried out using FRODO (Jones, 1978) on an Evans & Sutherland PS300.

2.9 Discussion.

The crystal structure of 212-842 is shown in Figure 2.4, the coordinates of which are shown in table 2.1a-b. Table 2.2 a-c lists bond lengths, bond angles and torsion angles. Figure 2.5 shows a stereodiagram of the molecular packing. The associated bond distances and angles are in good agreement with tabulated values (Allen, Kennard, Watson, Orpen & Taylor, 1987). While the structure has acceptably low temperature factors the three methoxy groups on C(3), C(26) and C(28) show some of the largest vibrational motions. This may be due to the fact that they are not involved in any bonding interactions. The malonic acid is a planar structure. The charge of the deprotonated carboxyl group C(33) is delocalised, both C-O bond lengths being almost equivalent. The other carboxyl group of C(36) is not delocalised, retaining it's hydroxyl hydrogen. The nitrogen N(17) is protonated probably as a result of the pH at which the crystal was grown in malonic acid and thus the drug has crystallised as an acid salt. Protonation of N(17) has not affected the chair conformation of the ring. The protonation of this nitrogen occurs in one other of the structures found in the database search, oxypertine (QUEST88 REFCODE = CAGXIR). This compound is an α-adrenoreceptor antagonist. The interaction of α-ligands with their receptor are believed to involve the quaternary Nitrogen (Pullman, Coubeils, Courriere & Gervois, 1972)

so the capacity of 212-842 to become protonated might be very important in its activity.

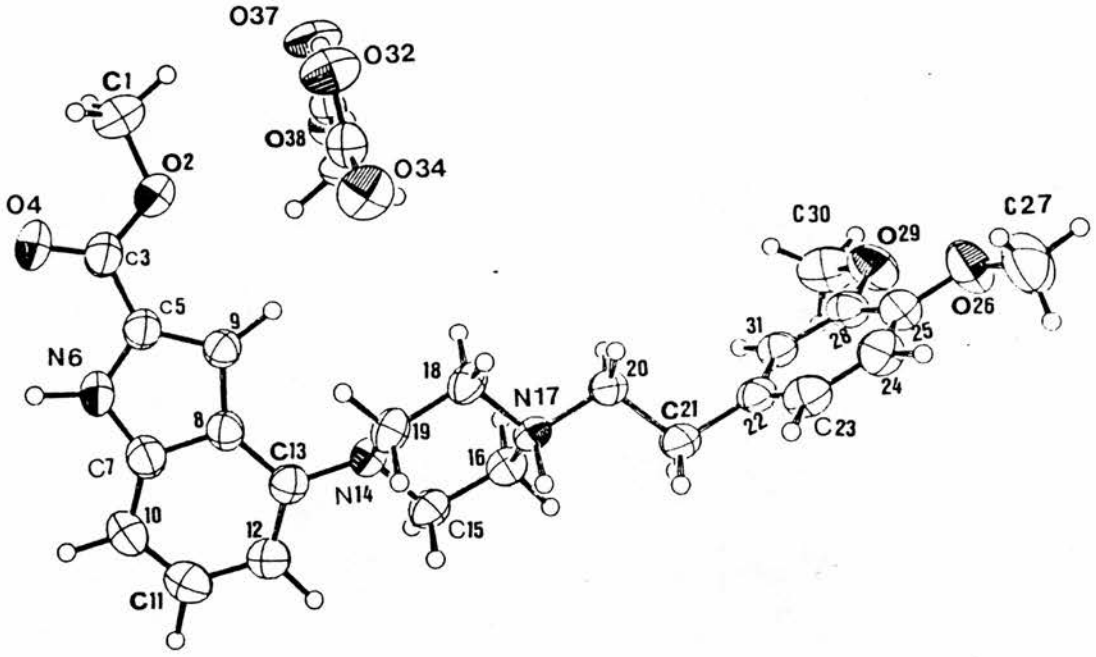


Figure 2.4. 212-842 crystal structure. Non-H atoms are shown as 50% probability thermal ellipsoides. The counterion derived from malonic acid is also shown.

Table 2.1(a). Fractional Coordinates of 212-842 non-H Atoms with Standard Deviations.

	x	y	z	Ueq
C(1)	0.41890(22)	0.7057(3)	0.21466(21)	0.0872(23)
O(2)	0.47039(12)	0.58603(19)	0.21420(11)	0.0700(12)
C(3)	0.48796(18)	0.5375(3)	0.14376(16)	0.0609(17)
O(4)	0.46035(16)	0.58525(23)	0.08348(12)	0.0967(16)
C(5)	0.54527(16)	0.42339(25)	0.14833(14)	0.0527(15)
N(6)	0.57831(14)	0.37430(22)	0.08010(12)	0.0608(14)
C(7)	0.63299(17)	0.2725(3)	0.09936(14)	0.0567(16)
C(8)	0.63240(15)	0.25507(23)	0.18212(13)	0.0464(14)
C(9)	0.57664(15)	0.35347(23)	0.21177(14)	0.0468(14)
C(10)	0.68388(19)	0.1970(3)	0.05046(16)	0.0714(19)
C(11)	0.73504(20)	0.1050(3)	0.08685(17)	0.0739(20)
C(12)	0.73554(17)	0.0832(3)	0.16912(15)	0.0609(17)
C(13)	0.68405(14)	0.15535(23)	0.21758(14)	0.0467(14)
N(14)	0.68094(11)	0.13706(19)	0.30040(11)	0.0437(11)
C(15)	0.59838(14)	0.0960(3)	0.32774(14)	0.0493(14)
C(16)	0.59411(14)	0.11314(25)	0.41530(14)	0.0488(14)
N(17)	0.65961(10)	0.02893(17)	0.45588(10)	0.0384(10)
C(18)	0.74437(13)	0.06127(24)	0.42495(13)	0.0448(13)
C(19)	0.74496(14)	0.05047(25)	0.33625(14)	0.0485(14)
C(20)	0.65571(14)	0.04427(23)	0.54390(13)	0.0444(13)
C(21)	0.69855(16)	-0.0699(3)	0.58760(14)	0.0547(15)
C(22)	0.68467(14)	-0.06161(24)	0.67511(13)	0.0467(14)
C(23)	0.62916(15)	-0.1467(3)	0.71066(16)	0.0561(16)
C(24)	0.61641(17)	-0.1396(3)	0.79125(16)	0.0611(17)
C(25)	0.65838(16)	-0.0471(3)	0.83662(15)	0.0567(16)
O(26)	0.65026(14)	-0.03072(23)	0.91597(11)	0.0839(15)
C(27)	0.59321(25)	-0.1163(4)	0.95448(21)	0.106(3)
C(28)	0.71535(15)	0.04004(25)	0.80120(14)	0.0501(14)
O(29)	0.75471(12)	0.12864(20)	0.85167(11)	0.0728(13)
C(30)	0.81660(19)	0.2130(3)	0.82039(20)	0.0804(22)
C(31)	0.72752(14)	0.03372(24)	0.72140(14)	0.0472(14)
O(32)	0.52564(11)	0.80055(17)	0.42820(12)	0.0683(12)
C(33)	0.58566(15)	0.72211(23)	0.41586(13)	0.0430(13)
O(34)	0.66015(10)	0.75696(16)	0.41640(10)	0.0574(11)
C(35)	0.56603(14)	0.57594(23)	0.39863(15)	0.0481(14)
C(36)	0.47729(16)	0.5280(3)	0.40671(14)	0.0526(15)
O(37)	0.42071(11)	0.61983(20)	0.42093(13)	0.0735(13)
O(38)	0.45923(12)	0.41054(19)	0.40071(12)	0.0757(14)

Table 2.1(b). Fractional Coordinates of 212-842 Hydrogen Atoms

	x	y	z	Ueq
H(11)	0.3990	0.7226	0.2742	0.0600
H(12)	0.3645	0.6923	0.1763	0.0600
H(13)	0.4546	0.7910	0.1951	0.0600
H(6N)	0.5637	0.4099	0.0212	0.0600
H(91)	0.5620	0.3697	0.2728	0.0600
H(101)	0.6828	0.2105	-0.0127	0.0600
H(111)	0.7768	0.0466	0.0514	0.0600
H(121)	0.7771	0.0085	0.1946	0.0600
H(151)	0.5506	0.1571	0.2993	0.0600
H(152)	0.5878	-0.0081	0.3129	0.0600
H(161)	0.5329	0.0828	0.4345	0.0600
H(162)	0.6045	0.2173	0.4302	0.0600
H(171)	0.6475	-0.0756	0.4432	0.0600
H(181)	0.7611	0.1623	0.4420	0.0600
H(182)	0.7897	-0.0081	0.4500	0.0600
H(191)	0.7327	-0.0522	0.3192	0.0600
H(192)	0.8058	0.0803	0.3158	0.0600
H(201)	0.6860	0.1372	0.5608	0.0600
H(202)	0.5907	0.0474	0.5602	0.0600
H(211)	0.7650	-0.0656	0.5774	0.0600
H(212)	0.6736	-0.1638	0.5656	0.0600
H(231)	0.5951	-0.2198	0.6755	0.0600
H(241)	0.5730	-0.2077	0.8182	0.0600
H(271)	0.5957	-0.0909	1.0163	0.0600
H(272)	0.6093	-0.2208	0.9473	0.0600
H(273)	0.5306	-0.0989	0.9309	0.0600
H(301)	0.8406	0.2737	0.8686	0.0600
H(302)	0.7859	0.2766	0.7770	0.0600
H(303)	0.8678	0.1603	0.7937	0.0600
H(311)	0.7703	0.1026	0.6943	0.0600
H(351)	0.6048	0.5165	0.4383	0.0600
H(352)	0.5835	0.5568	0.3386	0.0600
H(37O)	0.4520	0.7061	0.4264	0.0600

Table 2.2(a). 212-842 bond Lengths(A) with standard deviations

C(1)	-H(11)	1.080(5)	C(19)	-H(191)	1.080(3)
C(1)	-H(12)	1.080(5)	C(19)	-H(192)	1.080(3)
C(1)	-H(13)	1.080(5)	C(20)	-H(201)	1.080(3)
C(1)	- O(2)	1.448(4)	C(20)	-H(202)	1.080(3)
O(2)	- C(3)	1.326(3)	C(20)	-C(21)	1.512(3)
C(3)	- O(4)	1.202(4)	C(21)	-H(211)	1.080(4)
C(3)	- C(5)	1.460(4)	C(21)	-H(212)	1.080(4)
C(5)	- N(6)	1.375(3)	C(21)	-C(22)	1.509(3)
C(5)	- C(9)	1.367(3)	C(22)	-C(23)	1.377(3)
N(6)	-H(6N)	1.080(3)	C(22)	-C(31)	1.400(3)
N(6)	- C(7)	1.372(3)	C(23)	-H(231)	1.080(4)
C(7)	- C(8)	1.416(3)	C(23)	-C(24)	1.390(4)
C(7)	-C(10)	1.396(4)	C(24)	-H(241)	1.080(4)
C(8)	- C(9)	1.424(3)	C(24)	-C(25)	1.366(4)
C(8)	-C(13)	1.416(3)	C(25)	-O(26)	1.366(3)
C(9)	-H(91)	1.080(3)	C(25)	-C(28)	1.403(4)
C(10)	-H(101)	1.080(4)	O(26)	-C(27)	1.419(5)
C(10)	-C(11)	1.364(4)	C(27)	-H(271)	1.080(6)
C(11)	-H(111)	1.080(4)	C(27)	-H(272)	1.080(6)
C(11)	-C(12)	1.413(4)	C(27)	-H(273)	1.080(6)
C(12)	-H(121)	1.080(4)	C(28)	-O(29)	1.371(3)
C(12)	-C(13)	1.379(4)	C(28)	-C(31)	1.375(3)
C(13)	-N(14)	1.419(3)	O(29)	-C(30)	1.411(4)
N(14)	-C(15)	1.466(3)	C(30)	-H(301)	1.080(5)
N(14)	-C(19)	1.458(3)	C(30)	-H(302)	1.080(5)
C(15)	-H(151)	1.080(3)	C(30)	-H(303)	1.080(5)
C(15)	-H(152)	1.080(3)	C(31)	-H(311)	1.080(3)
C(15)	-C(16)	1.499(3)	O(32)	-C(33)	1.258(3)
C(16)	-H(161)	1.080(3)	C(33)	-O(34)	1.238(3)
C(16)	-H(162)	1.080(3)	C(33)	-C(35)	1.517(3)
C(16)	-N(17)	1.495(3)	C(35)	-H(351)	1.080(3)
N(17)	-H(171)	1.0800(24)	C(35)	-H(352)	1.080(3)
N(17)	-C(18)	1.498(3)	C(35)	-C(36)	1.504(4)
N(17)	-C(20)	1.505(3)	C(36)	-O(37)	1.312(3)
C(18)	-H(181)	1.080(3)	C(36)	-O(38)	1.210(3)
C(18)	-H(182)	1.080(3)	O(37)	-H(370)	0.997(20)
C(18)	-C(19)	1.510(3)			

Table 2.2(b). 212-842 angles(degrees) with standard deviations

H(11) - C(1) -H(12)	109.5(4)	H(152)-C(15) -C(16)	109.28(24)
H(11) - C(1) -H(13)	109.5(4)	C(15) -C(16) -H(161)	109.30(24)
H(11) - C(1) - O(2)	108.3(3)	C(15) -C(16) -H(162)	109.31(24)
H(12) - C(1) -H(13)	109.5(4)	C(15) -C(16) -N(17)	110.12(19)
H(12) - C(1) - O(2)	110.0(3)	H(161)-C(16) -H(162)	109.5(3)
H(13) - C(1) - O(2)	110.1(3)	H(161)-C(16) -N(17)	109.31(23)
C(1) - O(2) - C(3)	116.00(23)	H(162)-C(16) -N(17)	109.32(23)
O(2) - C(3) - O(4)	122.8(3)	C(16) -N(17) -H(171)	109.31(18)
O(2) - C(3) - C(5)	112.51(23)	C(16) -N(17) -C(18)	110.14(16)
O(4) - C(3) - C(5)	124.7(3)	C(16) -N(17) -C(20)	110.76(16)
C(3) - C(5) - N(6)	119.05(22)	H(171)-N(17) -C(18)	107.20(17)
C(3) - C(5) - C(9)	130.95(24)	H(171)-N(17) -C(20)	106.52(17)
N(6) - C(5) - C(9)	109.97(22)	C(18) -N(17) -C(20)	112.74(16)
C(5) - N(6) -H(6N)	125.74(24)	N(17) -C(18) -H(181)	109.02(21)
C(5) - N(6) - C(7)	108.52(21)	N(17) -C(18) -H(182)	109.01(21)
H(6N) - N(6) - C(7)	125.74(24)	N(17) -C(18) -C(19)	111.26(18)
N(6) - C(7) - C(8)	107.88(21)	H(181)-C(18) -H(182)	109.5(3)
N(6) - C(7) -C(10)	129.27(24)	H(181)-C(18) -C(19)	109.03(22)
C(8) - C(7) -C(10)	122.83(24)	H(182)-C(18) -C(19)	109.03(22)
C(7) - C(8) - C(9)	106.61(21)	N(14) -C(19) -C(18)	110.60(19)
C(7) - C(8) -C(13)	119.32(21)	N(14) -C(19) -H(191)	109.19(23)
C(9) - C(8) -C(13)	134.05(22)	N(14) -C(19) -H(192)	109.19(23)
C(5) - C(9) - C(8)	107.00(21)	C(18) -C(19) -H(191)	109.19(23)
C(5) - C(9) -H(91)	126.5(3)	C(18) -C(19) -H(192)	109.18(23)
C(8) - C(9) -H(91)	126.5(3)	H(191)-C(19) -H(192)	109.5(3)
C(7) -C(10) -H(101)	121.8(3)	N(17) -C(20) -H(201)	108.74(22)
C(7) -C(10) -C(11)	116.3(3)	N(17) -C(20) -H(202)	108.73(22)
H(101)-C(10) -C(11)	121.8(3)	N(17) -C(20) -C(21)	112.40(18)
C(10) -C(11) -H(111)	118.7(3)	H(201)-C(20) -H(202)	109.5(3)
C(10) -C(11) -C(12)	122.6(3)	H(201)-C(20) -C(21)	108.75(23)
H(111)-C(11) -C(12)	118.7(3)	H(202)-C(20) -C(21)	108.73(23)
C(11) -C(12) -H(121)	119.3(3)	C(20) -C(21) -H(211)	108.97(25)
C(11) -C(12) -C(13)	121.36(25)	C(20) -C(21) -H(212)	108.98(25)
H(121)-C(12) -C(13)	119.3(3)	C(20) -C(21) -C(22)	111.46(20)
C(8) -C(13) -C(12)	117.49(22)	H(211)-C(21) -H(212)	109.5(3)
C(8) -C(13) -N(14)	118.46(20)	H(211)-C(21) -C(22)	108.96(25)
C(12) -C(13) -N(14)	124.04(21)	H(212)-C(21) -C(22)	108.98(25)
C(13) -N(14) -C(15)	113.80(18)	C(21) -C(22) -C(23)	120.64(21)
C(13) -N(14) -C(19)	116.58(18)	C(21) -C(22) -C(31)	120.50(21)
C(15) -N(14) -C(19)	109.15(18)	C(23) -C(22) -C(31)	118.86(22)
N(14) -C(15) -H(151)	109.28(23)	C(22) -C(23) -H(231)	119.6(3)
N(14) -C(15) -H(152)	109.29(23)	C(22) -C(23) -C(24)	120.78(24)
N(14) -C(15) -C(16)	110.23(19)	H(231)-C(23) -C(24)	119.6(3)
H(151)-C(15) -H(152)	109.5(3)	C(23) -C(24) -H(241)	119.7(3)
H(151)-C(15) -C(16)	109.27(24)	C(23) -C(24) -C(25)	120.5(3)

H(241)	-C(24)	-C(25)	119.8(3)
C(24)	-C(25)	-O(26)	125.39(24)
C(24)	-C(25)	-C(28)	119.33(24)
O(26)	-C(25)	-C(28)	115.27(23)
C(25)	-O(26)	-C(27)	117.40(25)
O(26)	-C(27)	-H(271)	107.4(4)
O(26)	-C(27)	-H(272)	111.8(4)
O(26)	-C(27)	-H(273)	109.3(4)
H(271)	-C(27)	-H(272)	109.5(5)
H(271)	-C(27)	-H(273)	109.5(5)
H(272)	-C(27)	-H(273)	109.5(5)
C(25)	-C(28)	-O(29)	114.98(22)
C(25)	-C(28)	-C(31)	120.19(23)
O(29)	-C(28)	-C(31)	124.83(22)
C(28)	-O(29)	-C(30)	117.55(21)
O(29)	-C(30)	-H(301)	106.7(3)
O(29)	-C(30)	-H(302)	107.3(3)
O(29)	-C(30)	-H(303)	114.3(3)
H(301)	-C(30)	-H(302)	109.5(4)
H(301)	-C(30)	-H(303)	109.5(4)
H(302)	-C(30)	-H(303)	109.5(4)
C(22)	-C(31)	-C(28)	120.32(22)
C(22)	-C(31)	-H(311)	119.84(25)
C(28)	-C(31)	-H(311)	119.8(3)
O(32)	-C(33)	-O(34)	123.99(22)
O(32)	-C(33)	-C(35)	118.35(21)
O(34)	-C(33)	-C(35)	117.67(20)
C(33)	-C(35)	-H(351)	107.15(23)
C(33)	-C(35)	-H(352)	107.16(23)
C(33)	-C(35)	-C(36)	118.54(20)
H(351)	-C(35)	-H(352)	109.5(3)
H(351)	-C(35)	-C(36)	107.15(24)
H(352)	-C(35)	-C(36)	107.17(24)
C(35)	-C(36)	-O(37)	116.66(22)
C(35)	-C(36)	-O(38)	121.51(23)
O(37)	-C(36)	-O(38)	121.82(24)
C(36)	-O(37)	-H(370)	105.9(12)

Table 2.2(c). 212-842 torsion angles(degrees) with standard deviations

H(11) - C(1) - O(2) - C(3)	168.6(3)	C(11) -C(12) -C(13) - C(8)	-1.9(4)
H(12) - C(1) - O(2) - C(3)	49.0(4)	C(11) -C(12) -C(13) -N(14)	179.24(24)
H(13) - C(1) - O(2) - C(3)	-71.7(4)	H(121)-C(12) -C(13) - C(8)	178.1(3)
C(1) - O(2) - C(3) - O(4)	-3.1(4)	H(121)-C(12) -C(13) -N(14)	-0.8(4)
C(1) - O(2) - C(3) - C(5)	175.54(23)	C(8) -C(13) -N(14) -C(15)	61.8(3)
O(2) - C(3) - C(5) - N(6)	-171.05(22)	C(8) -C(13) -N(14) -C(19)	-169.73(20)
O(2) - C(3) - C(5) - C(9)	6.6(4)	C(12) -C(13) -N(14) -C(15)	-119.3(3)
O(4) - C(3) - C(5) - N(6)	7.6(4)	C(12) -C(13) -N(14) -C(19)	9.2(3)
O(4) - C(3) - C(5) - C(9)	-174.8(3)	C(13) -N(14) -C(15) -H(151)	-45.4(3)
C(3) - C(5) - N(6) -H(6N)	-2.9(4)	C(13) -N(14) -C(15) -H(152)	74.4(3)
C(3) - C(5) - N(6) - C(7)	177.10(23)	C(13) -N(14) -C(15) -C(16)	-165.46(19)
C(9) - C(5) - N(6) -H(6N)	179.0(3)	C(19) -N(14) -C(15) -H(151)	-177.50(23)
C(9) - C(5) - N(6) - C(7)	-1.0(3)	C(19) -N(14) -C(15) -H(152)	-57.7(3)
C(3) - C(5) - C(9) - C(8)	-177.8(3)	C(19) -N(14) -C(15) -C(16)	62.40(24)
C(3) - C(5) - C(9) -H(91)	2.2(5)	C(13) -N(14) -C(19) -C(18)	169.16(19)
N(6) - C(5) - C(9) - C(8)	0.0(3)	C(13) -N(14) -C(19) -H(191)	-70.7(3)
N(6) - C(5) - C(9) -H(91)	180.0(3)	C(13) -N(14) -C(19) -H(192)	49.0(3)
C(5) - N(6) - C(7) - C(8)	1.6(3)	C(15) -N(14) -C(19) -C(18)	-60.19(23)
C(5) - N(6) - C(7) -C(10)	-176.9(3)	C(15) -N(14) -C(19) -H(191)	60.0(3)
H(6N) - N(6) - C(7) - C(8)	-178.4(3)	C(15) -N(14) -C(19) -H(192)	179.65(23)
H(6N) - N(6) - C(7) -C(10)	3.1(5)	N(14) -C(15) -C(16) -H(161)	179.84(23)
N(6) - C(7) - C(8) - C(9)	-1.6(3)	N(14) -C(15) -C(16) -H(162)	60.0(3)
N(6) - C(7) - C(8) -C(13)	-179.92(21)	N(14) -C(15) -C(16) -N(17)	-60.07(24)
C(10) - C(7) - C(8) - C(9)	177.07(25)	H(151)-C(15) -C(16) -H(161)	59.7(3)
C(10) - C(7) - C(8) -C(13)	-1.3(4)	H(151)-C(15) -C(16) -H(162)	-60.1(3)
N(6) - C(7) -C(10) -H(101)	-2.7(5)	H(151)-C(15) -C(16) -N(17)	179.82(22)
N(6) - C(7) -C(10) -C(11)	177.3(3)	H(152)-C(15) -C(16) -H(161)	-60.0(3)
C(8) - C(7) -C(10) -H(101)	179.0(3)	H(152)-C(15) -C(16) -H(162)	-179.8(3)
C(8) - C(7) -C(10) -C(11)	-1.0(4)	H(152)-C(15) -C(16) -N(17)	60.0(3)
C(7) - C(8) - C(9) - C(5)	1.0(3)	C(15) -C(16) -N(17) -H(171)	-62.52(24)
C(7) - C(8) - C(9) -H(91)	-179.0(3)	C(15) -C(16) -N(17) -C(18)	55.02(23)
C(13) - C(8) - C(9) - C(5)	179.0(3)	C(15) -C(16) -N(17) -C(20)	-179.59(18)
C(13) - C(8) - C(9) -H(91)	-1.0(5)	H(161)-C(16) -N(17) -H(171)	57.6(3)
C(7) - C(8) -C(13) -C(12)	2.7(3)	H(161)-C(16) -N(17) -C(18)	175.10(22)
C(7) - C(8) -C(13) -N(14)	-178.38(21)	H(161)-C(16) -N(17) -C(20)	-59.5(3)
C(9) - C(8) -C(13) -C(12)	-175.1(3)	H(162)-C(16) -N(17) -H(171)	177.37(23)
C(9) - C(8) -C(13) -N(14)	3.8(4)	H(162)-C(16) -N(17) -C(18)	-65.1(3)
C(7) -C(10) -C(11) -H(111)	-178.1(3)	H(162)-C(16) -N(17) -C(20)	60.3(3)
C(7) -C(10) -C(11) -C(12)	1.9(4)	C(16) -N(17) -C(18) -H(181)	66.9(3)
H(101)-C(10) -C(11) -H(111)	1.9(5)	C(16) -N(17) -C(18) -H(182)	-173.61(21)
H(101)-C(10) -C(11) -C(12)	-178.1(3)	C(16) -N(17) -C(18) -C(19)	-53.34(23)
C(10) -C(11) -C(12) -H(121)	179.6(3)	H(171)-N(17) -C(18) -H(181)	-174.23(22)
C(10) -C(11) -C(12) -C(13)	-0.5(4)	H(171)-N(17) -C(18) -H(182)	-54.8(3)
H(111)-C(11) -C(12) -H(121)	-0.4(5)	H(171)-N(17) -C(18) -C(19)	65.49(23)
H(111)-C(11) -C(12) -C(13)	179.5(3)	C(20) -N(17) -C(18) -H(181)	-57.3(3)

```
C(20) -N(17) -C(18) -H(182)   62.1( 3)
C(20) -N(17) -C(18) -C(19) -177.60(18)
C(16) -N(17) -C(20) -H(201) -78.92(25)
C(16) -N(17) -C(20) -H(202)   40.2( 3)
C(16) -N(17) -C(20) -C(21)  160.62(19)
H(171)-N(17) -C(20) -H(201) 162.30(22)
H(171)-N(17) -C(20) -H(202) -78.6( 3)
H(171)-N(17) -C(20) -C(21)   41.85(24)
C(18) -N(17) -C(20) -H(201)   45.0( 3)
C(18) -N(17) -C(20) -H(202) 164.12(21)
C(18) -N(17) -C(20) -C(21)  -75.45(22)
N(17) -C(18) -C(19) -N(14)   56.39(24)
N(17) -C(18) -C(19) -H(191) -63.8( 3)
N(17) -C(18) -C(19) -H(192) 176.56(22)
H(181)-C(18) -C(19) -N(14)  -63.9( 3)
H(181)-C(18) -C(19) -H(191) 175.9( 3)
H(181)-C(18) -C(19) -H(192)  56.3( 3)
H(182)-C(18) -C(19) -N(14)  176.65(22)
H(182)-C(18) -C(19) -H(191)  56.5( 3)
H(182)-C(18) -C(19) -H(192) -63.2( 3)
N(17) -C(20) -C(21) -H(211)  66.8( 3)
N(17) -C(20) -C(21) -H(212) -52.6( 3)
N(17) -C(20) -C(21) -C(22) -172.93(18)
H(201)-C(20) -C(21) -H(211) -53.7( 3)
H(201)-C(20) -C(21) -H(212)-173.1( 3)
H(201)-C(20) -C(21) -C(22)   66.6( 3)
H(202)-C(20) -C(21) -H(211)-172.8( 3)
H(202)-C(20) -C(21) -H(212)  67.8( 3)
H(202)-C(20) -C(21) -C(22)  -52.5( 3)
C(20) -C(21) -C(22) -C(23)  104.8( 3)
C(20) -C(21) -C(22) -C(31)  -74.6( 3)
H(211)-C(21) -C(22) -C(23) -134.9( 3)
H(211)-C(21) -C(22) -C(31)   45.7( 3)
H(212)-C(21) -C(22) -C(23)  -15.5( 4)
H(212)-C(21) -C(22) -C(31)  165.1( 3)
C(21) -C(22) -C(23) -H(231)  -0.2( 4)
C(21) -C(22) -C(23) -C(24)  179.83(23)
C(31) -C(22) -C(23) -H(231) 179.2( 3)
C(31) -C(22) -C(23) -C(24)   -0.8( 4)
C(21) -C(22) -C(31) -C(28) -179.34(22)
C(21) -C(22) -C(31) -H(311)   0.7( 4)
C(23) -C(22) -C(31) -C(28)    1.3( 4)
C(23) -C(22) -C(31) -H(311)-178.7( 3)
C(22) -C(23) -C(24) -H(241)-179.5( 3)
C(22) -C(23) -C(24) -C(25)    0.5( 4)
H(231)-C(23) -C(24) -H(241)   0.5( 5)
H(231)-C(23) -C(24) -C(25) -179.5( 3)
C(23) -C(24) -C(25) -O(26)  179.5( 3)
C(23) -C(24) -C(25) -C(28)   -0.7( 4)
H(241)-C(24) -C(25) -O(26)   -0.5( 5)
H(241)-C(24) -C(25) -C(28)  179.3( 3)
C(24) -C(25) -O(26) -C(27)   -0.3( 4)
C(28) -C(25) -O(26) -C(27)  179.8( 3)
C(24) -C(25) -C(28) -O(29) -179.62(23)
C(24) -C(25) -C(28) -C(31)    1.2( 4)
O(26) -C(25) -C(28) -O(29)    0.2( 3)
O(26) -C(25) -C(28) -C(31) -179.03(23)
C(25) -O(26) -C(27) -H(271) 179.0( 3)
C(25) -O(26) -C(27) -H(272)  58.9( 5)
C(25) -O(26) -C(27) -H(273) -62.4( 4)
C(25) -C(28) -O(29) -C(30)  176.06(23)
C(31) -C(28) -O(29) -C(30)   -4.8( 4)
C(25) -C(28) -C(31) -C(22)   -1.5( 4)
C(25) -C(28) -C(31) -H(311) 178.5( 3)
O(29) -C(28) -C(31) -C(22)  179.40(22)
O(29) -C(28) -C(31) -H(311)  -0.6( 4)
C(28) -O(29) -C(30) -H(301)-179.3( 3)
C(28) -O(29) -C(30) -H(302)  63.5( 4)
C(28) -O(29) -C(30) -H(303) -58.1( 4)
O(32) -C(33) -C(35) -H(351)-128.6( 3)
O(32) -C(33) -C(35) -H(352) 114.0( 3)
O(32) -C(33) -C(35) -C(36)   -7.3( 3)
O(34) -C(33) -C(35) -H(351)  51.9( 3)
O(34) -C(33) -C(35) -H(352) -65.6( 3)
O(34) -C(33) -C(35) -C(36)  173.15(21)
C(33) -C(35) -C(36) -O(37)    6.4( 3)
C(33) -C(35) -C(36) -O(38) -173.27(23)
H(351)-C(35) -C(36) -O(37)  127.7( 3)
H(351)-C(35) -C(36) -O(38)  -52.0( 3)
H(352)-C(35) -C(36) -O(37) -114.9( 3)
H(352)-C(35) -C(36) -O(38)   65.4( 3)
C(35) -C(36) -O(37) -H(370)  -3.1(12)
O(38) -C(36) -O(37) -H(370) 176.6(12)
```

Table 2.3. 212-842 anisotropic Vibration Parameters with Standard Deviations

	U11	U22	U33	U23	U13	U12
C(1)	0.0965(25)	0.0768(22)	0.0882(23)	0.0023(19)	-0.0026(19)	0.0392(19)
O(2)	0.0844(13)	0.0671(13)	0.0580(11)	0.0041(10)	-0.0082(10)	0.0278(11)
C(3)	0.0731(18)	0.0582(17)	0.0511(16)	0.0051(14)	-0.0110(14)	0.0070(14)
O(4)	0.1382(20)	0.0962(17)	0.0550(12)	0.0094(12)	-0.0229(12)	0.0509(16)
C(5)	0.0642(16)	0.0477(15)	0.0461(14)	0.0018(12)	-0.0082(12)	0.0015(13)
N(6)	0.0816(16)	0.0586(14)	0.0420(12)	0.0055(10)	-0.0064(11)	0.0077(12)
C(7)	0.0705(17)	0.0521(16)	0.0474(14)	-0.0002(12)	-0.0013(13)	-0.0015(14)
C(8)	0.0551(14)	0.0421(14)	0.0420(13)	0.0005(11)	-0.0028(11)	-0.0046(11)
C(9)	0.0554(14)	0.0413(13)	0.0436(13)	0.0003(11)	-0.0053(11)	-0.0014(11)
C(10)	0.0961(22)	0.0733(20)	0.0450(15)	-0.0023(14)	0.0083(15)	0.0077(18)
C(11)	0.0878(22)	0.0721(20)	0.0621(18)	-0.0066(16)	0.0185(16)	0.0152(17)
C(12)	0.0688(17)	0.0562(16)	0.0578(16)	0.0007(13)	0.0080(13)	0.0084(14)
C(13)	0.0488(14)	0.0415(13)	0.0497(14)	0.0008(11)	0.0000(11)	-0.0043(11)
N(14)	0.0397(10)	0.0430(11)	0.0484(11)	0.0068(9)	-0.0004(8)	0.0006(9)
C(15)	0.0397(13)	0.0518(15)	0.0563(15)	0.0116(12)	-0.0046(11)	-0.0021(11)
C(16)	0.0363(12)	0.0531(15)	0.0570(15)	0.0121(12)	0.0002(11)	0.0081(11)
N(17)	0.0361(10)	0.0309(10)	0.0479(11)	0.0002(8)	-0.0033(8)	0.0000(8)
C(18)	0.0332(12)	0.0466(14)	0.0547(14)	0.0008(11)	-0.0009(10)	-0.0004(10)
C(19)	0.0405(13)	0.0487(14)	0.0563(15)	0.0037(12)	0.0004(11)	0.0041(11)
C(20)	0.0477(13)	0.0401(13)	0.0454(13)	-0.0033(11)	-0.0001(10)	0.0057(11)
C(21)	0.0623(16)	0.0529(15)	0.0488(14)	-0.0058(12)	-0.0071(12)	0.0200(13)
C(22)	0.0457(13)	0.0459(14)	0.0482(13)	-0.0045(11)	-0.0071(11)	0.0143(11)
C(23)	0.0516(15)	0.0488(15)	0.0677(17)	-0.0069(13)	-0.0085(13)	-0.0004(12)
C(24)	0.0595(16)	0.0586(17)	0.0653(17)	0.0028(14)	0.0033(13)	-0.0067(14)
C(25)	0.0599(16)	0.0597(17)	0.0505(15)	-0.0001(13)	0.0031(12)	0.0046(14)
O(26)	0.1032(16)	0.0968(16)	0.0519(11)	-0.0029(11)	0.0167(11)	-0.0117(13)
C(27)	0.1227(31)	0.1217(33)	0.0748(22)	0.0127(22)	0.0346(22)	-0.0091(26)
C(28)	0.0504(14)	0.0489(15)	0.0509(15)	-0.0086(12)	-0.0054(11)	0.0022(12)
O(29)	0.0807(13)	0.0746(13)	0.0630(12)	-0.0210(10)	-0.0013(10)	-0.0155(11)
C(30)	0.0718(20)	0.0676(20)	0.1014(25)	-0.0211(18)	-0.0106(18)	-0.0108(17)
C(31)	0.0439(13)	0.0449(14)	0.0526(14)	-0.0015(12)	-0.0009(11)	0.0058(11)
O(32)	0.0617(11)	0.0387(10)	0.1050(15)	-0.0065(10)	0.0156(10)	0.0054(9)
C(33)	0.0507(15)	0.0377(13)	0.0404(13)	-0.0015(10)	0.0001(10)	-0.0043(12)
O(34)	0.0526(11)	0.0451(10)	0.0745(12)	-0.0076(9)	0.0009(9)	-0.0128(8)
C(35)	0.0450(13)	0.0373(13)	0.0619(15)	-0.0090(12)	0.0033(11)	-0.0054(11)
C(36)	0.0525(15)	0.0541(17)	0.0509(15)	0.0002(12)	-0.0045(12)	-0.0104(13)
O(37)	0.0439(10)	0.0644(13)	0.1121(16)	0.0003(12)	0.0080(10)	-0.0031(10)
O(38)	0.0762(13)	0.0534(12)	0.0971(15)	-0.0042(11)	-0.0070(11)	-0.0275(10)

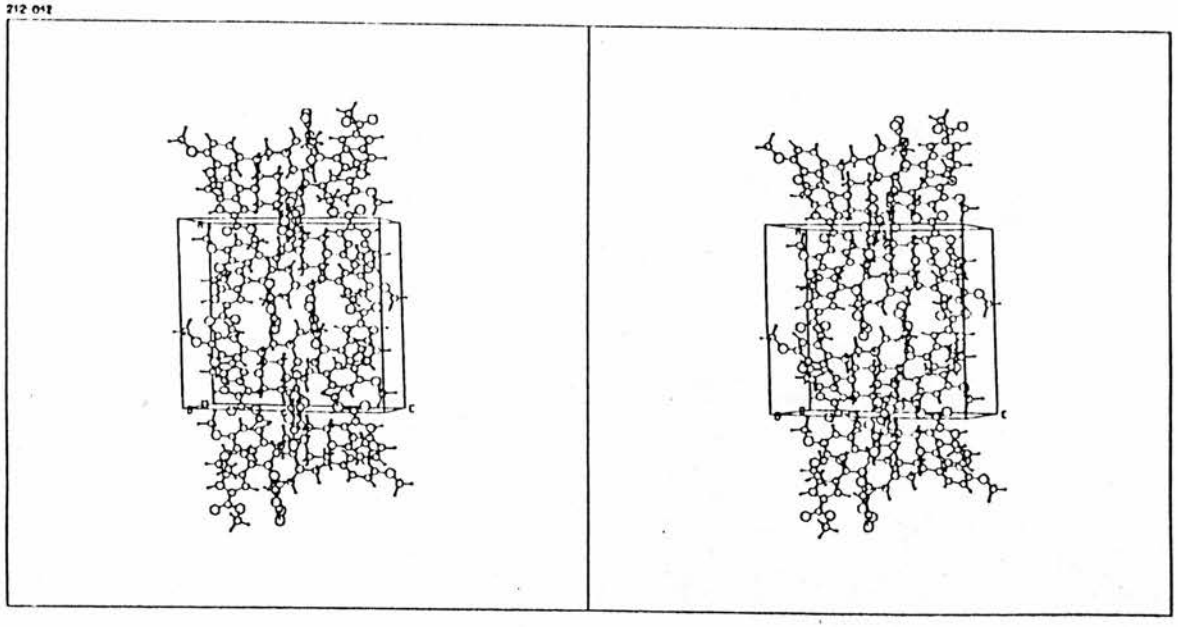


Figure 2.5. Stereopacking diagram of 212-842, b-axis projection.

2.9.1 Hydrogen bonding.

There are 3 strong hydrogen bonds between molecules in the crystal structure of 212-842. The intermolecular distances of these and the symmetry operators of the groups involved are shown in table 2.4. The malonic acid forms an intramolecular hydrogen bond between the carboxyl oxygen O(32) and the hydroxyl group O(37)H. The strongest hydrogen bond is that between the two charged species, namely 212-842 and the malonic acid. There are two other strong hydrogen bonds, these are between symmetry related molecules of 212-842. As shown in table 2.4 donor H(6N) is bonded to the acceptor O(4) of a molecule related to it by symmetry (1-x,1-y,-z). However the H(6N) of this symmetry related molecule is also hydrogen bonded to O(4) of the molecule at (x,y,z). Therefore there are two identical centrosymmetrically related hydrogen bonds represented by that particular entry in table 2.4.

2.9.2 Crystal packing.

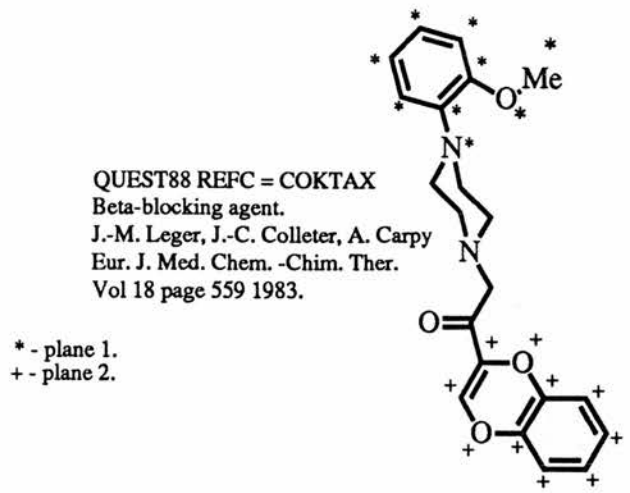
The crystal packing interactions are of two types, namely hydrogen bonding and hydrophobic ring stacking. The hydrogen bonding has already been described above. The majority of van der Waals contacts are between molecules related by the symmetry operation (3/2-x,1/2+y,1-z) and occur at both ends of the piperazine ring. Thus the substituted benzene ring stacks to the indolyl ring of the symmetry related molecule and vice versa at the other end.

2.9.3 Comparison with related structures.

The crystallographic coordinates of two related structures shown in figure 2.6 were obtained from the Cambridge Structural Database for use in the comparison. Initially the structures were superimposed

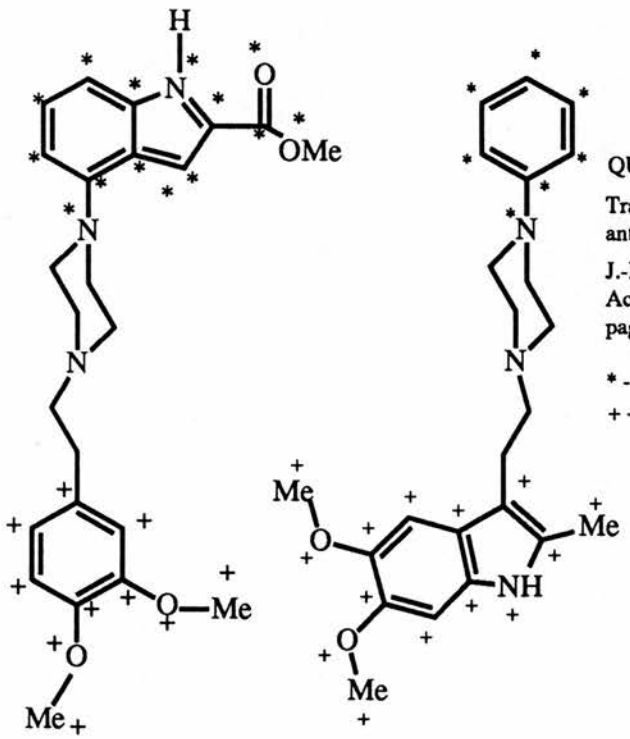
Table 2.4 The major hydrogen bonding in 212-842 showing distances (A) and the symmetry operators of the acceptors involved.

H(171)	-	O(34)	(X,-1+Y,Z)	1.7430(23)
H(6N)	-	O(4)	(1-X,1-Y,-Z)	1.809(3)
H(171)	-	O(32)	(X,-1+Y,Z)	2.3120(25)



QUEST88 REFC = COKTAX
Beta-blocking agent.
J.-M. Leger, J.-C. Colleter, A. Carpy
Eur. J. Med. Chem. -Chim. Ther.
Vol 18 page 559 1983.

* - plane 1.
+ - plane 2.



QUEST88 REFC = CAGXIR

Tranquiliser , Alpha-adrenoreceptor antagonist.

J.-M. Leger, M.Saux, A.Carpy
Acta Cryst.,C (cr.Str.Comm.) Vol 39 page1428 1983.

* - plane 1.
+ - plane 2.

212-842

* - plane 1
+ - plane 2

Figure 2.6 Structure of 212-842 showing the related molecules found in the database search. The atoms defining the planes used for the interplanar angle and C.O.G/C.O.G. distances are indicated.

manually using FRODO on the Evans & Sutherland PS300. The fragment used in the database search was used as the basis for superposition. All structures possessed a chair conformation for the piperazine ring. The equivalent substituent aromatic groups and the two planes which define them, for all three molecules, are shown in figure 2.6. All three structures superimposed particularly well and it was clear that the equivalent substituent aromatic groups were approximately coplanar and that these planes were almost at right angles to one another as shown in plate 2.1. The superposition was carried out more objectively using CALC. It was found that all the interplanar angles fell within 12° of each other (Table 2.5). The mean value for the interplanar angle being $77.86 \pm 6^{\circ}$. The centres of gravity of these planar groups at either end of the piperazine ring were calculated using CALC, for all the structures and their separations compared (Table 2.5). The centres of gravity of the planar moieties at either end of the piperazine ring have a mean separation of 11.15 ± 0.21Å. Further analysis of similar compounds may prove the interplanar angle and the C.O.G./C.O.G distance to be important structural features related to their activity. Additional work may be necessary to show these structural features are retained when these compounds are in solution but that is beyond the scope of the work here. It may be that there are other structural features that are common to this group of similar compounds that may be understood once the structure of their receptors has been elucidated.

Plate 2.1 Stereo picture of the superposition of 212-842 with two related molecules. Superposition was performed manually using FRODO as described in the experimental section.

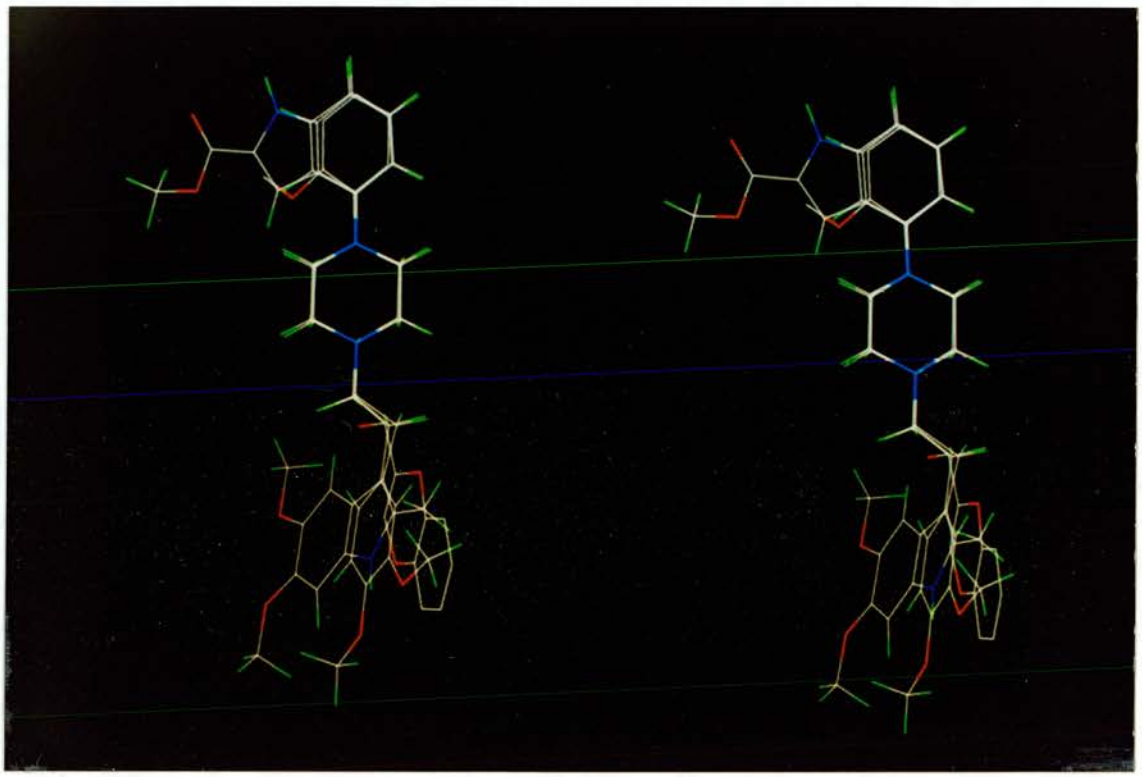


Table 2.5
Structural features identified with 212-842 that are shared by two other molecules of similar biological activity.

Compound	Interplanar Angle (°)	COG/COG Distance (Å)*
212-842	83.47	11.36
CAGXIR **	78.31	11.16
COKTAX **	71.81	10.92

* - Centre of gravity (COG) of planar moiety in each compound.

** - QUEST88 REFCODE

Chapter 3 Purification of ovine pro-phospholipase A2

Chapter 3

3.1.Introduction

A necessary requirement for any structural studies on a protein is that the protein be available to a high level of purity and in relatively large quantities. Typically from 20 mg up to gram quantities may be required. Such large quantities are required for crystallisation trials and for the crystal structure determination itself. Thus the protein has to be purified on a large scale from some tissue in which the protein is abundant. More recently, some genetically engineered organism in which the protein has been cloned and over expressed can be used.

The non-venom extracellular phospholipase-A2 enzymes (phosphatide acyl-hydrolase, EC 3.1.1.4), due to their relative abundance and their stability, have been purified to homogeneity in a number of species. The extracellular enzyme is synthesised in the pancreas as a pro-enzyme which is activated by trypsin in the stomach by removal of the N-terminal heptapeptide. Pancreatic tissue of ox (Rimon & Shapiro, 1959), man (Magee *et al.*, 1962), horse (van den Bosch *et al.*, 1965) and rat (Hanahan, 1962) have been used as the enzyme source. The purification of these extracellular enzymes has been facilitated by the observation that they are stable at elevated temperatures and acidic pH. The method described here is based on that used to purify Bovine PlA-2 (Dutihl *et al.*, 1975) and equine PlA-2 (Zurini, M. personal communication). Since the amino acid sequences of these extracellular PlA-2's are very homologous, the general methods used in their purification are also very similar. Only the chromatographic properties of the different enzymes require alterations to the conditions used in the purification. Typically 200-300 mgs of pure

enzyme can be obtained per kg of pancreas. The aim of the purification was to obtain sufficient quantities of the zymogen and the active enzyme from sheep pancreas for structural studies. No crystal structure has yet been obtained for any of the extracellular pro-phospholipase A2's, although the mechanism of activation is well understood.

3.2 Materials and Method.

3.2.1 Materials.

Bovine Trypsin (Type III) and phenylmethylsulphonylfluoride (PMSF) were obtained from Sigma. All other reagents were of reagent-grade purity. Diethylaminoethyl cellulose (DE-52) and Carboxymethyl cellulose (CM-52) were purchased from Whatman and treated according to the manufacturers instructions. The stirred cell membranes (YM10) were purchased from Amicon and treated according to the manufacturers instructions prior to use.

3.2.2 Protein measurement.

Protein concentration was measured either by the Lowry procedure (Lowry *et al.*, 1951), using crystallised bovine serum albumin (Sigma) as standard, or spectrophotometrically at 280nm using the molar absorbance coefficient $A_{1\%1cm} = 13$ determined previously for the bovine pro-enzyme (Dutihl *et al.*, 1975).

3.2.3 Gel electrophoresis.

The electrophoretic procedures used were those of Laemmli (Laemmli, 1970). Standards used were low molecular weight markers (Sigma) and Bovine phospholipase A2 (Sigma). All gels

were 12.5% acrylamide unless otherwise specified.

3.2.4 Tissue homogenisation and heat treatment.

Ovine pancreas was harvested from freshly slaughtered animals and stored over ice during transportation to the laboratory. Subsequent operations were carried out in a cold room at 4°C. The pancreas were defatted and connective tissue removed before being cut into small pieces. The tissue was then homogenised in 3000 ml/kg of ice-cold 150 mM NaCl using a Waring blendor. Homogenisation was carried out in three bursts of 3mins in duration. A 2 minute period between each burst was allowed for cooling purposes. The pH of the homogenate was adjusted to 4.0 using 6M HCl under efficient stirring. The homogenate was then heated rapidly under efficient stirring to 70°C and the temperature held for 3 minutes. The homogenate was then rapidly cooled to 4°C using an acetone/dry-ice mixture. This was also performed under efficient stirring. The homogenate was then centrifuged for 30 mins at 20 000g in an ultracentrifuge at 4°C and the supernatant harvested. The supernatant was freed from fat by filtration through filter paper after addition of 5-10 g of Hyflo supercel/l of supernatant.

3.2.5 Ammonium sulphate fractionation.

The filtrate was brought to pH 7.0 with concentrated ammonia solution. Solid ammonium sulphate (333.1 g/l of filtrate, 0.4S, where S refers to fractional saturation) was slowly added under efficient stirring. The solution was stirred at 4°C for 2 hours. The subsequent precipitate was removed by centrifugation in a ultracentrifuge for 30mins at 20 000g at 4°C. From the clear supernatant the enzyme was precipitated by a second addition of solid ammonium sulphate (129 g/l

of filtrate, 0.65S) which was slowly added under efficient stirring. The solution was again stirred at 4°C for 2 hours before the subsequent precipitate was removed by centrifugation for 30 mins at 20 000g and the supernatant discarded.

The precipitate was redissolved in a minimal amount of distilled water using a combination of gentle agitation and sonication. To this was added 5 vol-% of 0.1M solution of phenylmethylsulphonyl -fluoride (PMSF), in dry isopropanol. The resulting protein solution containing PMSF was then dialysed overnight in running water at 4°C. The small precipitate formed was removed by centrifugation for 20 minutes at 15 900g at 4°C. After further addition of 2 vol-% 0.1M PMSF solution, the solution (approximately 250mls) was lyophilised.

3.2.6 Ion exchange chromatography.

Two chromatographic steps were used to purify the zymogen. The first was an anion exchange step and this was followed by a cation exchange step. The lyophilised material (about 6g) was dissolved in 150 mls of 5mM Tris-HCl buffer pH 8.0 at 4°C. The subsequent chromatographic steps were carried out at 4°C.

DE-52 Chromatography.

The first step was carried out using an anion exchange column of dimensions 1.5 cm x 12.5 cm. The matrix was equilibrated in 5mM Tris-HCl buffer pH 8.0 in the cold room at 4°C. 150mls of protein solution was applied to the column using a peristaltic pump at a flow rate of 1.5 mls/min. This was followed by 1000 mls of buffer. After the non absorbed proteins had been eluted, a linear salt gradient of 1.5l of 5mM Tris-HCl buffer from 0 to 0.4M NaCl was used to elute the

remaining proteins from the column. The samples were collected in 10 ml fractions. The absorbance at 280nm was measured on a Philips PU8700 spectrophotometer. Only fractions containing significant protein as determined from their A_{280nm} were assayed for enzyme activity. Those fractions containing enzyme activity after trypsin activation were pooled and dialysed against 2 x 5 litres of 5mM acetate/acetic acid buffer pH 6.0 at 4°C to remove salt and exchange the buffer. The solution was then concentrated to a final volume of 25 mls using a stirred cell containing a YM10 membrane (Amicon,) which has a molecular weight cut-off of ten thousand. This was not necessary since the next stage was an ion exchange column, but it was done to speed up the purification as there was such a large volume of dialysed material.

CM-52 Chromatography.

The secondstep was carried out using a cation exchange column of dimensions 1.5 x 12.5 cm. The matrix was equilibrated in 5mM sodium acetate/acetic acid buffer, pH 6.0 at 4°C. A volume of 25 mls of protein solution (approximately 10mg/ml) was applied to the column using a peristaltic pump at a flow rate of 1.5 mls/min. After the non-absorbed proteins had passed through the column, a linear gradient of 200 mls 5mM acetate buffer containing 0 to 0.4M NaCl was used to elute the remaining proteins from the column. The non absorbed proteins were collected in fractions of 10 mls whereas the proteins eluted with the gradient were collected in 5 ml fractions. The protein concentration of the fractions was determined by the absorbance at 280nm using a Pye Unicam SP1800 spectrophotometer.

3.2.7 Trypsinisation of sample for assay.

Trypsinisation was based on the the method of Abita & Lazdunski, (1972). Samples (0.1ml) were added to the activation buffer (0.9ml) prior to the addition of trypsin (1μg). The activation was carried out at 4°C and pH 8. The activation buffer contained 10mM $CaCl_2$, 10mM Tris-HCl buffer pH 8. The final concentration of the activating trypsin was 1μg/ml and was added from a stock solution of 1mg/ml. The incubation volume was 1ml. The period of activation was 60 mins. Tryptic digestion of the enzyme was stopped by dilution (1:600) of this sample (50μl) in the assay mixture (30mls).

3.2.8 Enzyme Assays.

Phospholipase A2 activity was routinely measured potentiometrically in a pH-stat Radiometer TTT1C autotitrator equipped with a SBU-1a syringe and a combined glass-calomel electrode at 40°C using 0.1M NaOH as the titrant. The action of the enzyme on the phospholipids is to liberate a fatty acid into the medium. This acid is titrated by alkali and the pH of the medium is maintained at pH 8. The volume of the autoburette was 2.5mls and the consumption of alkali was automatically recorded. As substrate, an aqueous emulsion of egg yolk was used, prepared by homogenizing one egg yolk in 100 ml of water. Per assay, 10 ml of the yolk emulsion was diluted to 30 ml with stock $CaCl_2$ and sodium deoxycholate solutions, the final concentration of sodium deoxycholate and $CaCl_2$ in the assay mixture being $2.7x10^{-3}$M and $6.0x10^{-3}$M, respectively. Nitrogen was bubbled through the assay mix for 15 minutes and then the pH was adjusted to 8.0 prior to the addition of enzyme. The enzyme had been activated as described previously. Basal activity was measured for 3½ minutes for each assay

mixture before the addition of enzyme. The volume of activated sample used in the enzyme assay was 50μl. The consumption of alkali was measured over a seven minute period after the addition of the sample. Activity is expressed as the uptake of alkali in μequiv/min and was calculated by measuring the activity and subtracting basal activity. Thus one unit of activity liberates one micromole of titratable fatty acid per minute in the egg yolk suspension at pH 8.0 and 37°C under the specified conditions. Specific activity is given by the number of μequiv of alkali consumed per min, per mg of protein. The assay system was initially tested using known amounts of firstly 0.1 M HCl and then bovine phospholipase A2 (Sigma: 48 units/mg protein) and shown to be over 95% accurate in both cases.

3.2.9 Bulk activation of pro-phospholipase by trypsin.

The typical activation was carried out at pH 8.0 at 4°C, in 50 mM Tris-HCl, 10mM $CaCl_2$ for one hour. Trypsin solution was freshly made up at a concentration of 1 mg/ml in the same buffer. The necessary volume of trypsin solution was added to the protein solution giving a relative molar ratio of trypsin to pro-phospholipase in the activation mixture of 1:100. After 30mins of incubation a second quantity of fresh trypsin was added. The activation was terminated by lowering the pH of the solution to 6.

3.3 Results.

The results of the initial purification are shown in table 3.1. The total amount of defatted pancreas used was 1.35kg. This produced approximately 4.35l of crude homogenate. Subsequently this was reduced to 3.3l after heat treatment and removal of cell debris and

Table 3.1. Purification of pro-PLA-2 from Ovine pancreas.
Specific activities were measured after the activation of the enzyme with trypsin, in the egg yolk assay as described in the methods, and are expressed in μequiv. acid released min-1 mg-1 of protein. Protein concentration measured at 280nm.

Step	Volume (ml)	Total * zymogen Activity μequiv.min-1	Specific * Activity of zymogen μequiv.min-1 .mg-1	Yield (%)	Total Protein (mg)
1. Crude Homogenate **	4250	___ +	___ +	___ +	___
2. Heat Treatment, filtration & centrifugation	3290	56 792	2.2	100	25800
3. Ammonium Sulphate fractionation	105	35 779	5.8	63	6175
4. DE-52 chromatography	60	25 556	54.1	45	472
5. CM-52 chromatography	40	20 445	87	36	235

* Determined as Phospholipase A2 activity after activation of the zymogen with Trypsin.

** 1.35 kilograms of defatted pancreas was homogenised as described in the protocol.

+ No activity could be detected on account of the endogenous Trypsin inhibitors.

precipitated protein. After rapid cooling there was a large floating layer of fat. After its removal, the final quantity of the pale yellowliquid was 3.295 litres. These initial stages of the purification were also followed by gel electrophoresis using various samples from the specified stages, Plate 3.1. Approximately 6g of lyophilised material was obtained after the ammonium sulphate fractionation. On redissolving, this lyophilised material has a slight yellow colour to it. The elution profile of ovine prophospholipase from the DE 52 column is shown in Fig. 3.1, which yielded one peak containing activity which eluted at approximately 65mM NaCl. The first and main peak eluted by the NaCl contains this activity, however it is not clear that this is a single peak and may be a combination of unresolved peaks. The material containing the yellow colour did not absorb to the exchanger and was obtained with the flow-through along with the non absorbed proteins. The flow-through contained no detectable enzymic activity after tryptic activation. From a protein estimation of the pooled fractions after dialysis it was estimated that there was 295mg of prophospholipase A2. This extensively dialysed material was applied to the CM 52 column. The elution profile from this is shown in Fig 3.2. The small peak associated with the flow-through did not contain any detectable activity after tryptic activation. A single peak was eluted at approximately 240mM NaCl. These fractions possessed associated activity after activation by trypsin. Electrophoresis of both a fraction from this peak and the same fraction after trypsin activation are shown in Plate 3.2, showing the slight molecular weight shift associated with activation. The two samples have the apparent molecular weight of M_r 14 400 for the proenzyme (from amino acid analysis $M_r \approx 14\,570$) and M_r 13 800 for the

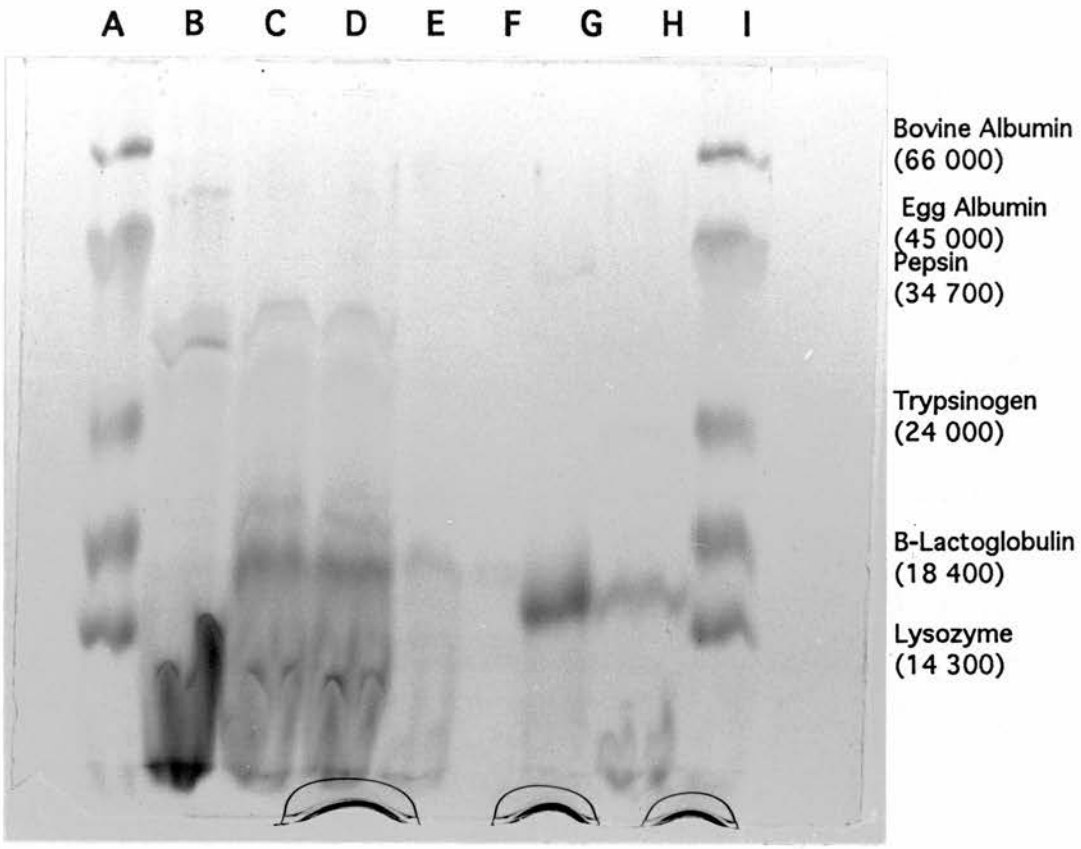


Plate 3.1

SDS-PAGE of samples from the purification of Ovine pro-PlA-2. Low molecular weight markers (Sigma) are shown in lanes A and I. The soluble component of the crude homogenate is shown in B, C is this fraction after heat treatment at pH 4. Lane D shows the soluble fraction C after removal of the floating fat layer. Lane E shows the soluble proteins at 0.4s and F shows the soluble proteins at 0.65s. Lane G contains purified ovine pro-PlA2 and lane H contains Bovine PlA2 (Sigma).

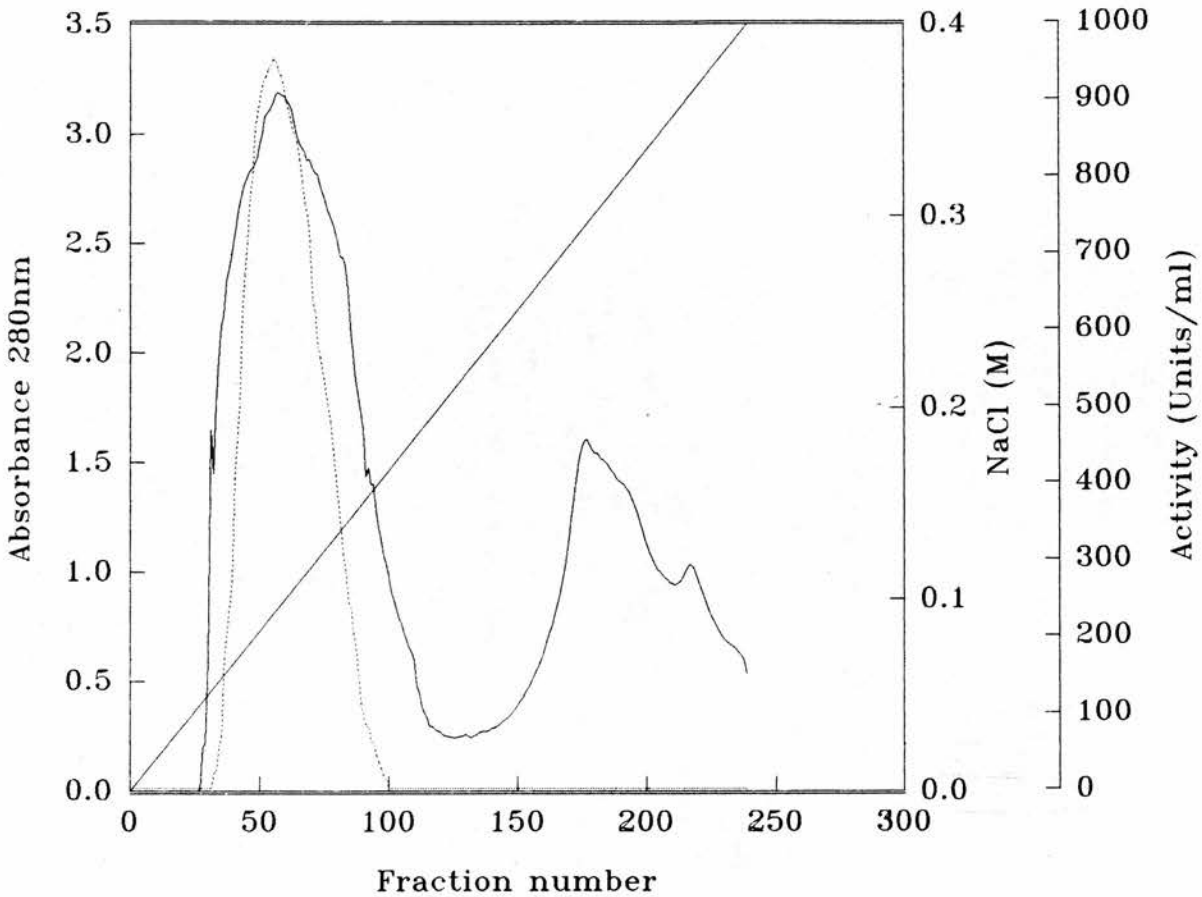


Figure 3.1. Elution profile of ovine pro-phospholipase A2 from a DE52 column (1.5 cm x 50 cm) at pH 8.0. ___, absorbance at 280nm; ---, enzyme units/ml after activation by Trypsin. The peak associated with the non absorbed proteins has been omitted for clarity.

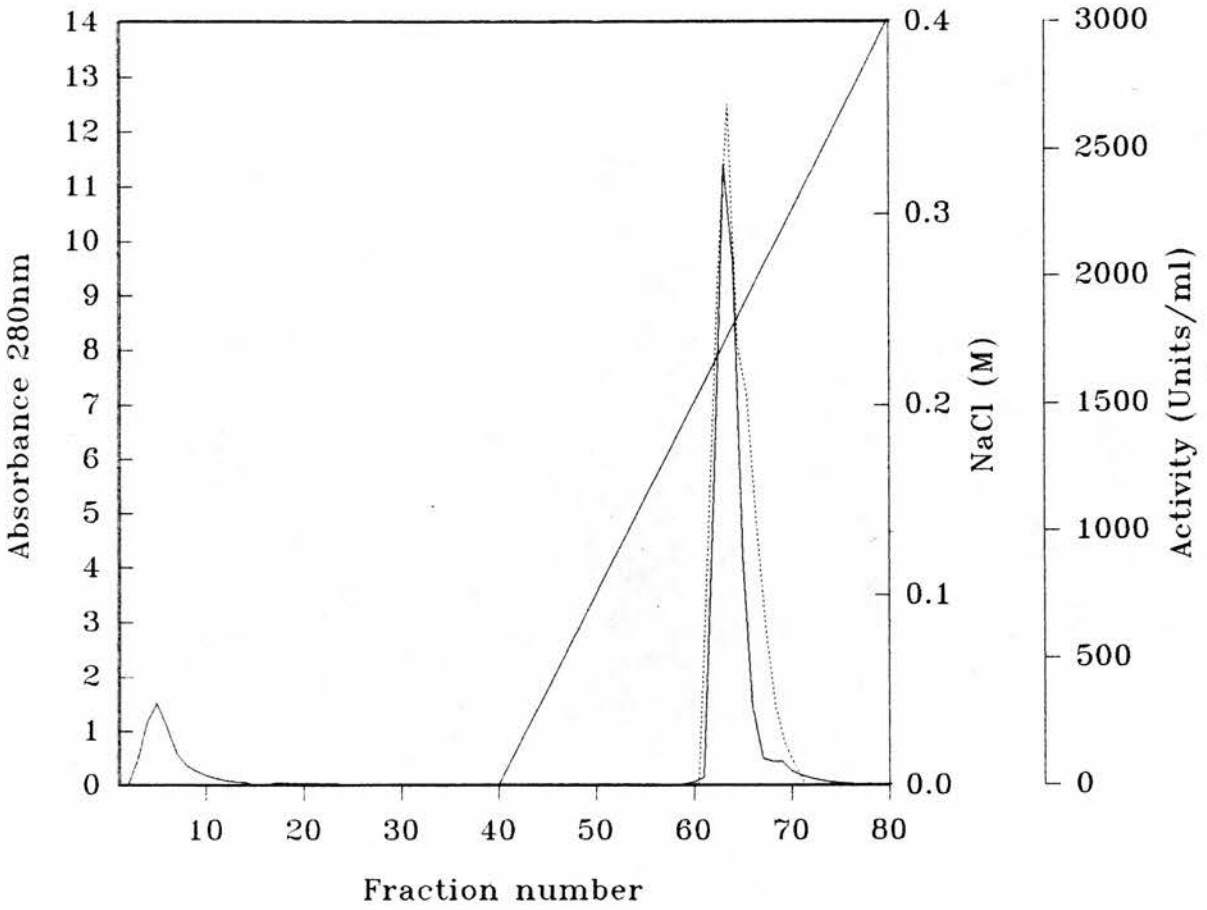


Figure 3.2. Elution profile of ovine pro-phospholipase A2 from a CM52 column (1.5 cm x 12.5 cm) at pH6.0. ___, absorbance at 280nm; ---, enzyme units/ml after activation by Trypsin. The gradient applied was 0-0.4 M NaCl in acetate buffer at pH 6.0.

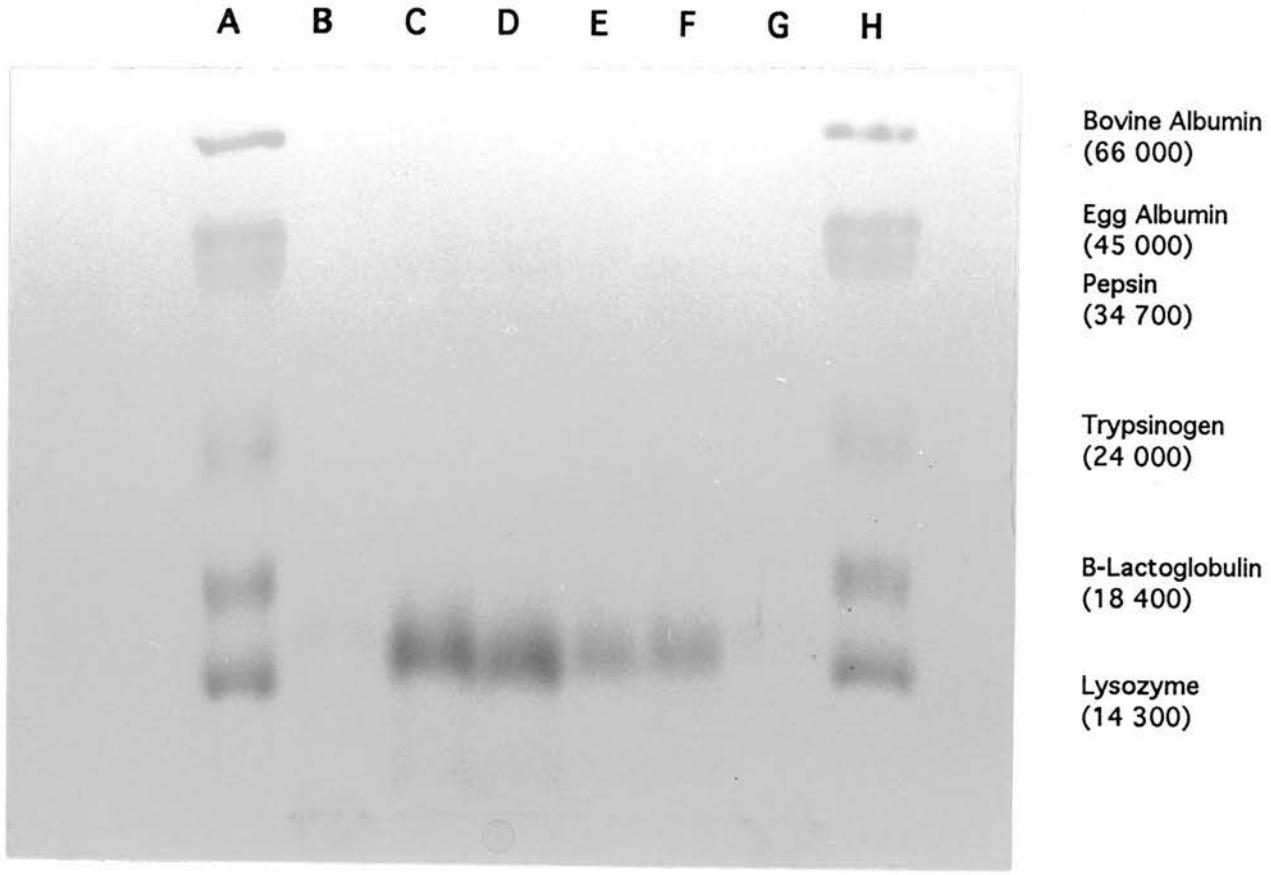


Plate 3.2
SDS-PAGE of purified ovine pro-PlA-2 before and after activation by trypsin. The activation was carried out as described in the methods. Lanes A and H contain low molecular weight markers (Sigma). Lanes B and G contain Bovine PlA2 (Sigma). Different loadings of ovine pro-PlA2 are shown in lanes C and F and different loadings of the activated enzyme are shown in lanes D and E.

active form of the enzyme (from amino acid analysis $M_r \approx 13\ 820$). The zymogen fractions were treated with PMSF and freed from salt by extensive dialysis at 4°C against distilled water. The yield after lyophilisation was 235 mg (= 79.6 % of column charge). The total recovery of zymogen was 89 %; however, a few fractions having a somewhat lower specific activity were discarded. The zymogen preparation as shown by gel electrophoresis (Plate 3.2) was at least 95% pure.

3.4 Discussion.

The use of freshly slaughtered pancreas for the isolation of the zymogen is crucially important. It has been shown that even after only two weeks in the frozen state, enzyme activity can be detected in porcine pancreas (Rimon and Shapiro, 1959) which has been attributed to activation of a fraction of the zymogen. Despite the presence of trypsinogen and trypsin in the pancreas, the need for the addition of anti-trypsin inhibitors during the initial homogenisation stages was not required. The pancreas also contains more than adequate supplies of endogenous trypsin inhibitors (Fritz *et al.*, 1967). However these are only effective during the initial stages of the purification. These natural inhibitors are either so effective or so abundant that no enzymic activity was detected in the crude homogenate after tryptic activation. All the steps after the ammonium sulphate fractionation were carried out in the presence of PMSF. This regime proved effective since none of the samples obtained prior to the salt precipitation contained a measurable activity before activation by trypsin. It is clear that the ovine enzyme is stable to heating at 70°C for 3 minutes at low pH, since no dramatic loss of activity was seen. Indeed it is

clear this is quite an effective purification step. Other workers have noted that the enzyme is stable at these high temperatures, even up to 97°C for a bacterial source (Hayaishi & Kornberg, 1954). The heat stability has been attributed to the presence of seven disulphides, which in such a relatively small protein provides structural stability. The structural stability conferred to the enzyme can be more easily understood from the locations of these bonds in the three dimensional structure itself. This is shown for the bovine enzyme (figure 1.4) in Chapter 1.

One of the consequences of this heat treatment, which has been previously identified, is the formation of pyro-glutamic acid at the N-terminus. This has been shown in the purification of the equine zymogen (Zurini, M. personal communication) and in the porcine zymogen (de Haas *et al.*, 1968). It has not been shown that any other residues of the zymogen have undergone rearrangements at these temperatures. From the gel, Plate 3.1, it is clear that there is very little loss of material with the removal of the floating layer of fat. There is undoubtedly some loss of material but this is not a major problem. This small loss is due to the association of the zymogen with the fatty material. Unfortunately this is unavoidable but both the rapid cooling to a low temperature and the low pH help to minimise this association. Work on porcine phospholipase A2 has shown that if the pH is raised to pH 7 before filtration almost all the enzyme was bound to the insoluble material (de Haas *et al.*, 1968). This material can be recovered from the insoluble material by treatment with sodium deoxycholate and solvent extraction with n-butanol to remove the lipids. This was not necessary in this case, since a substantial fraction of the zymogen remained soluble.

Despite being an excellent purification step (Plate 3.1) there is a quite substantial loss of activity after the ammonium sulphate fractionation (37% from Table 3.1). This suggests that the fractionation limits used 0.4s-0.65s (used for bovine and equine purification) are not optimal for the ovine purification. From the electrophoresis of these fractions (Plate 3.1) it is clear that a significant fraction of the enzyme is still soluble at 0.65s. There are also very few other proteins which are soluble at this concentration of ammonium sulphate. This upper concentration would therefore need to be raised for a more optimal purification of the ovine enzyme. This would not greatly affect the overall purification of the enzyme but would significantly improve the yield. The observation that the enzyme is more soluble in ammonium sulphate solutions than the bovine and equine enzyme might be of relevance when it comes to crystallisation trials with the enzyme.

The first chromatographic step, anion exchange with DE-52 (Figure 3.1), produced only two main eluted peaks - the largest containing the activity. This indicates that the initial stages were indeed excellent purification steps. Indeed this elution profile is very similar to those obtained for both bovine (Dutihl *et al.*, 1975) and equine (Zurini, M. personal communication) forms of the zymogen. The presence of multiple forms of the ovine zymogen were not detected either by the assay system or during the chromatographic steps. However, the major peak of the elution profile of the DE 52 column suggests the presence of other unresolved peaks. Two prophospholipases have been found for the porcine zymogen (Nieuwenhuizen *et al.*, 1973). The difference between the two has been shown to be located in the activation peptide. The major form has the heptapeptide with the pyroglutamic N-

terminus whereas the minor form (4%) is shorter and contains only the three amino acids Ser-Ser-Arg. It is thought that this second form could be generated from the first by cleavage with elastase, since it can be produced *in vitro* this way. However, it may be a possibility that this minor form is also a naturally occuring zymogen. These forms have slightly different isoelectric points and consequently it would be expected that they would behave differently on chromatographic steps. However, from the elution profile of the CM52 chromatographic step, it is clear that there is only one form, Figure 3.2. The electrophoresis (Plate 3.2) also indicates a single species, but it is doubtful that two species differing by three amino acids could be resolved under the conditions used in the electrophoresis despite it being possible to detect a molecular weight shift on tryptic activation where there is a difference of seven amino acids. Also since ovine PLA-2 runs as a rather broad band it is not possible to detect multiple forms even if they were present. However, since trypsin activation yields the same active enzyme regardless of the sequence of the pro-peptide, the exact sequence of the activation peptide of the purified zymogen has not been determined. It has been shown for the porcine enzyme that the tryptic activation kinetics of these two prophospholipases are very similar (Nieuwenhuizen *et al.*, 1973). Thus the activity measurements carried out during the ovine zymogen purification are not affected by the possible existence of two zymogens. It must also be noted that the activity measurements were carried out using Ca^{2+} and deoxycholate concentrations that were optimal for porcine PlA-2 (Nieuwenhuizen *et al.*, 1974). It is known that the optimal conditions with respect to Ca^{2+} and deoxycholate concentrations for enzymes from different sources can be very

different (Figarella *et al.*, 1971;Evenberg *et al.*, 1977) in this type of assay. Every enzyme has its characteristic optimum for these two components. These concentrations not only affect the rate of hydrolysis but also affect the linearity of the reaction with time. Outside the optimal conditions the reaction rapidly slows down with time, an effect that also has negative influence on proportionality between enzyme concentration and velocity (Verheij *et al.*, 1981). The activities measured for the ovine enzyme are of a similar magnitude to those measured during the purification of the bovine enzyme (Dutilh *et al.*, 1975). The activity of bovine and the measured activity of the ovine enzyme are both considerably smaller than those obtained for the porcine form, which possess an activity an order of magnitude larger (de Haas *et al.*, 1968). However the ruminant bovine phospholipase has been shown to be as active as the porcine enzyme in kinetic experiments with dioctanoyllecithin (Dutilh *et al.*, 1975). The low activity of the ruminant enzyme in the egg yolk assay system (mixed micelles of phospholipid + bile salts) was therefore attributed to a weak interaction between the enzyme and the organised lipid-water interface. Similar kinetic experiments on the ovine enzyme have not been carried out using dioctanoyllecithin. It is possible that a weak interaction between the enzyme and the organised lipid-water interface may also occur. This would explain the relatively low activity of the ovine enzyme, another ruminant enzyme, in this assay system. Since enzyme activity was being used only as a diagnostic of PLA-2 presence, the relatively low enzymic activity in the egg yolk assay has not hindered the purification of the enzyme.

Chapter 4 Protein crystallisation

Chapter 4

4.1 Introduction

The purpose of this chapter is to give a brief introduction to the theory of the methods used in the production of protein crystals. Some of these methods are applied in the attempted crystallisation of salmon calcitonin (sCT), two derivatives of salmon calcitonin and also the co-crystallisation of equine phospholipase-A2 with an inhibitory compound.

The determination of three-dimensional protein structures, as described for the small molecules in chapter 2, using X-ray crystallography, is mainly limited by the ability to grow crystals. A crystal, whether it be of a simple salt or of a complicated protein, represents an ordered array of molecules extending in three dimensions. The crystallisation of macromolecules is a relatively ancient chemical procedure (Osborne, 1892), and has also been used as a criterion of purity in protein purification. The principles of crystallisation for any protein are little different to those of conventional small molecules. The primary differences in protein crystallisation are due to the nature of the protein molecules themselves and arise from the size, low symmetry, large solvent content and small binding energies of proteins in crystals. The latter causes the protein crystals to be soft and sensitive to small changes in external conditions. The solution behaviour of macromolecules is complex and rather unpredictable owing to their varied shapes, polyvalent surface character and dynamic properties. Macromolecules can demonstrate a number of distinct solubility minima. These depend on the nature of the electrolyte, the precipitant, their concentrations, the concentration of the macromolecule, pH,

temperature and a variety of other parameters.

The crystallisation of molecules from solution is a reversible equilibrium phenomenon, the specific parameters required for the crystallisation depend on the chemical and physical properties of the solvent and solute involved. However, only under conditions of supersaturation will this equilibrium be driven towards a state where the solute will be partitioned between a soluble and solid phase. The formation of this solid phase obviously results in loss of entropic energy. This is compensated for by the formation of many new stable chemical bonds. Thus the driving force in this equilibrium is the minimisation of the free energy, which can only be achieved by the ordering of the system in the solid phase - the crystal.

While the crystallisation of small molecules can be scientifically treated (Strikland-Constable, 1968), macromolecular crystallisation is more complex, less easily described and poorly understood. In macromolecular systems the equilibrium often leads to what is called an "amorphous precipitate". In energy terms this corresponds to one local minimum and frequently occurs when aggregation occurs too rapidly. Experience has shown that if supersaturation can be approached slowly it can afford the molecules sufficient time to orient themselves into a crystalline lattice. Thus the technique of macromolecular crystallisation is to bring the system very slowly to a state of minimum solubility and thus achieve a degree of supersaturation.

The approach to supersaturation can be achieved in basically two ways. In the most common case, the precipitant level is gradually increased until the solubility minimum is reached. A more subtle technique is to increase the precipitant concentration beyond the

normal precipitation point of the protein, but at a temperature or pH at which the protein is still soluble. The temperature or pH are then allowed to relax slowly to a desired value and produce supersaturation. There are three main categories of precipitants which are often used in crystallisation trials: salts, such as ammonium sulphate; organic solvents, such as ethanol and methylpentanediol; and the polyethylene glycols. The success of these materials to induce crystallisation probably reflects the fashion of crystallographers rather than some innate quality of the materials themselves. There is also a wide range of methods available to produce the desired supersaturation. These have been reviewed elsewhere (McPherson, 1982), but the most common methods include direct addition, interface diffusion, microdialysis, vapour diffusion and slow cooling. The most commonly used method is the vapour diffusion or hanging drop technique. In this technique a small droplet (6 to 40 μl), containing the macromolecule, a buffer and a precipitating agent is equilibrated against a reservoir containing a solution of the precipitant at a higher concentration than that of the droplet. This is achieved by placing the droplet on a glass slide which is then inverted over the well containing the reservoir solution. Equilibration proceeds by evaporation of volatile species (normally water) until the vapour pressures are the same in the reservoir and the droplet. As evaporation progresses from the droplet, supersaturation develops and can induce crystallisation.

While there are methods and precipitants that have often been proved to be successful, the crystallisation of a new protein is achieved by a very much trial-and-error approach. Thus protein crystallisation remains unpredictable and is still regarded as more of an art than a

science. Attempts have been made to provide a more scientific approach to protein crystallisation. Much work has centred on the theory of protein solubility. The effect of salts on solubility has been explained in terms of the electrostatic interaction between small ions and protein charges, or dipoles, by the Debye-Huckel theory (Edsell & Wyman, 1958). Hydrophobic interactions have also been studied and shown to be of importance in protein-protein and protein-ligand interactions (Kauzmann, 1959). More recent work has concentrated on the study of protein solubility and preferential solvent-protein interactions at high concentrations of co-solvent (Arakawa & Timasheff, 1982). Much of this work has been based on the results obtained for only a handful of abundant proteins such as ß-lactoglobulin (Treece *et al.*, 1964). The necessary parameters required for such work having been determined by exhaustive experimentation. Thus the application of a theoretical approach to the solubility of new and less abundant proteins is limited.

The process of crystallisation can be reduced to three distinct phases: (1) nucleation, (2) post-nucleation growth and (3) cessation of growth. The nucleation stage is a very important step in the production of crystals and represents the formation of ordered aggregates. Post-nucleation growth involves various parameters determining the rate of growth of the crystal. Cessation of growth is of importance in protein crystallisation because often crystals are very small. The factors that prevent the growth of crystals are poorly understood. The nucleation process has been the subject of the work of Kam *et al.*, (1987), using lysozyme as a model system. This work produced a theoretical model of the crystallisation process for understanding nucleation using quasi-elastic light scattering.

Crystallisation is described as a cooperative step-by-step addition of monomers, whereas amorphous precipitation is described as a non-cooperative polymerisation process. According to this theoretical model, in solutions leading to crystallisation, an abrupt change in aggregate size is predicted, whereas a nearly linear variation is expected for precipitating solutions. However, the conclusions of this model have not been generalised to many other systems (Baldwin *et al.*, 1986; Carter *et al.*, 1988). Quasi-elastic light scattering has also been used to study the suitability of solvent conditions (precipitating agent and buffer) for crystallisation assays (Mikol *et al.*, 1990). Using concanavalin A and lysosyme it was demonstrated that the concentration dependence of the translational diffusion coefficient could be correlated to the ability to crystallise. If the diffusional coefficient varies in a solvent concentration-dependent way, then the protein will not crystallise when the solution becomes supersaturated. In contrast solvents which do not produce variations in the diffusional coefficient under the same conditions will lead to crystallisation under supersaturated protein concentrations. The explanation for this is that under such conditions up to saturation, the proteins remain essentially monodisperse. This work been applied to the crystallisation of yeast aspartyl-tRNA synthetase (Mikol *et al.*, 1990b). Another, less theoretical, approach to improving the search for crystallisation conditions is the use of incomplete factorial experiments (Carter & Carter, 1979). The method attempts to reduce the number of experiments of the more traditional systematic approach to protein crystallisation. The factorial method utilises statistical treatment of "individual factors", the permutations of these are randomly assigned. The problem with this type of study is

that the results are semi-quantitative and are of little use when only amorphous precipitates are obtained.

The setting up of crystallisation trials is by no means trivial and is rather time consuming. A recent technical development for the hanging drop method has been an automated liquid handling system (Cox & Weber, 1987). The system is under computer control and consists of a pipetting station and multiport rotary valve. This can significantly increase the rate at which experiments are set up and facilitates the systematic studies of protein crystallisation.

The hanging drop method, which is a vapour diffusion technique, is the most commonly used method for protein crystallisation. A theoretical model has been recently proposed to describe the hanging drop method (Fowlis *et al.*, 1988). One of the main conclusions of this model was that the rate-limiting step is diffusion of water vapour across the air space separating the drop and the reservoir. Water equilibration rates in the hanging drop method have been studied with different precipitating agents (Mikol *et al.*, 1990). The rate has been shown to be significantly affected by two parameters namely temperature and initial drop volume, equilibration rates being slower at lower temperatures and larger initial drop volumes. This study indicated that the presence of protein in the drop does not affect the equilibration rate of the water to any significant extent. The other parameters which also affect the rate are the water vapour pressure of the reservoir, the nature of the precipitant and the dilution factor of the drop relative to the reservoir. The time for water equilibration has been shown to vary between about 25 hours and 25 days, depending on the experimental conditions.

However, despite these advances in experimentation and understanding

of the crystallisation process, the complete picture is far from clear. In the cases where proteins have failed to crystallise no attempts have been made at an explanation. Undoubtedly, understanding why a protein resists crystallisation is as important as understanding why it does.

4.2 Materials and methods

4.2.1 Salmon calcitonin

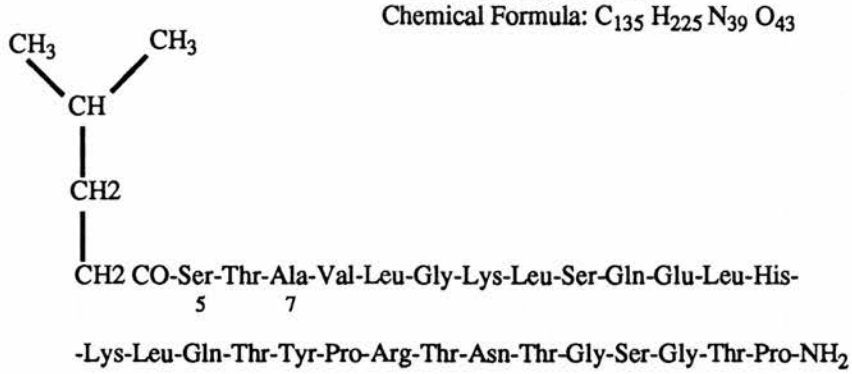
The salmon Calcitonin (sCT) was prepared by solid-phase chemistry, purified by preparative HPLC and was the generous gift of Sandoz AG. The peptide was obtained as a precipitate from tri-fluoroacetate. The counterion was replaced by acetate using FPLC on a G-25 column equilibrated in 50% Acetic Acid. The peak was collected and lyophilised before being used in crystallisation trials. The purity of the salmon Calcitonin was assessed by amino acid analysis using an LKB 44000 amino acid analyser. The peptide was hydrolysed at 110°C for 24 hours in an evacuated sealed tube containing 6 N HCl.

4.2.2 Salmon calcitonin derivatives

Two derivatives of salmon calcitonin were also used in crystallisation experiments. Both were prepared by solid phase chemistry in the same way as salmon calcitonin, purified by preparative HPLC and were also the generous gift of Sandoz AG. No further purification was necessary and no amino acid analysis was carried out on either derivative. The structures of the two derivatives are shown in Figure 4.1. The N^{α}-isocaproyl-[Ala7]-salmon calcitonin (5-32)-amide is a derivative that despite the loss of the

Derivative 1.

Molecular weight: 3082.5
Chemical Formula: $C_{135}\ H_{225}\ N_{39}\ O_{43}$



Derivative 2.

Molecular Weight: 3334.3
Chemical Formula: $C_{135}\ H_{223}\ I_2\ N_{39}\ O_{43}$

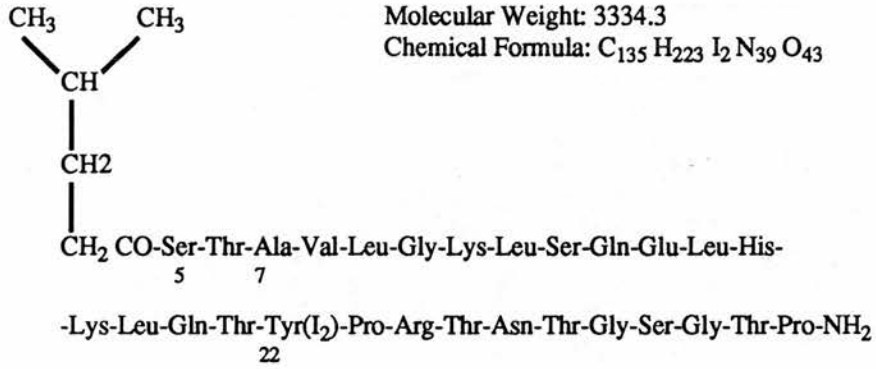


Figure 4.1.

The linear sequence of the two derivatives of salmon calcitonin used in the crystallisation trials. These were: Derivative 1 - N^{α}-Isocaproyl-[Ala^{7}]-salmon calcitonin-(5-32)-amide; Derivative 2 - N^{α}-Isocaproyl-[Ala^{7}, $Tyr(3,5\text{-}I_2)^{22}$]-salmon calcitonin-(5-32)-amide. Also shown are the Molecular Weights and Formulae of both.

amino terminal region still possesses some activity. The iodinated derivative N^{α}-isocaproyl-[Ala7,Tyr(3,5-I_2)22] salmon calcitonin (5-32) amide was available for use as a potential isomorphous derivative for the sCT analogue. The location of the iodine atoms in the unit cell would permit the phasing of the reflections. The iodinated derivative was kept in the dark and at -40°C prior to use.

4.2.3 Buffers

The buffers used in all the crystallisation trials and the range of pH's for which they were used are shown in Table 4.1. All buffers were 100mM and contained 0.02% sodium azide. The buffers were filtered prior to use through Durapore membrane filters, pore size 5μm (Millipore).

4.3 Crystallisation by slow cooling

This method was the initial method used for the attempted crystallisation of calcitonin, since this method has been successfully applied to the crystallisation of some other small polypeptide hormones such as glucagon (Sasaki *et al.*, 1975) and avian pancreatic polypeptide (Wood *et al.*, 1977). This method was applied only to the native hormone (sCT) and not the two derivatives.

General experimental set up

Buffered protein solutions were produced by addition of buffer at 70°C to lyophilised protein. These were then freed of any particulate material by centrifugation on a bench centrifuge for 5 minutes at 13000 rpm. A 5μl volume of this solution was placed in a melting point tube using a Hamilton syringe. The melting point tubes had been previously washed with an ethanol/water mixture and oven dried. To

Table 4.1 Buffers used in all the crystallisation trials showing the pH range for which they were used.

Buffer	pH range used
Glycine - HCl	3.0 - 3.8
Sodium Acetate - Acetic Acid	4.0 - 5.0
Sodium Cacodylate - HCl	5.2 - 6.8
Imidazole* -HCl	7.0 - 7.8
Tris** - HCl	8.0 - 8.5

* - Glyoxaline

** - Tris(hydroxymethyl)aminomethane

this was added an equal volume of precipitant at 70°C. The melting point tube was sealed and placed in a water bath at 70°C for at least 2 hours before being suspended in water at 70°C within a Dewar flask. The flask was sealed and further insulated by surrounding it with polyurethane foam. The boiling point tubes were left in this arrangement for five days after which time the water in the flask had reached room temperature. The tubes were examined under a light microscope for any signs of crystallisation. Only five pH's were studied in these experiments namely 4, 5, 6, 7, and 8. An initial experiment was carried out to determine the maximum peptide concentration which was not supersaturated in the absence of precipitating agents, under the experimental conditions. Subsequent experiments were performed to study the effects of a variety of different precipitants and effects of pH on the solubility of calcitonin under the experimental conditions.

4.3.1 Effect of calcitonin concentration

Stock solutions of calcitonin at each pH were produced by dissolving 5 mg of calcitonin in 50 μl of 0.1M buffer at 70°C. These were then spun for 5 minutes at 13000 rpm on a bench centrifuge. The stock solutions were thus 100 mg/ml. These were subsequently used to make further dilutions with buffer at 70°C to produce calcitonin solutions approximately 50, 25, 10 and 5 mg/ml. The melting point tubes containing these solutions were incubated as described previously.

4.3.2 Effect of some divalent metal ions

Calcitonin is known to affect calcium metabolism in osteoclasts (MacIntyre *et al.*, 1987) and transmembrane zinc transport in Leydig

cells (Chausmer *et al.*, 1989). Additionally the crystallisation of some small polypeptides, for example insulin (Adams *et al.*, 1969) and avian pancreatic polypeptide (Blundell *et al.*, 1981), has required the presence of a divalent metal ion. The possible requirement of a divalent ion in the nucleation process leading to crystallisation was investigated. Calcitonin was dissolved in buffer at 70°C to a final concentration of 50 mg/ml. To this was added an equal volume of a solution of either $ZnSO_4$ or $CaCl_2$ also at 70°C. The final concentrations of these salts were 14.6 mM, 29.2 mM, 43.8 and 100 mM. This represented a molar ratio of calcitonin:divalent ion of 1:1, 1:2, 1:3 and 1:6, the assumption being that calcitonin M_r= 3431.9. The final calcitonin concentration was 25 mg/ml.

4.3.3 Effect of some organic solvents

A number of organic solvents have been used in the crystallisation of small polypeptides, for example bovine growth hormone (Bell *et al.*, 1985). The solvents which were used in these slow cooling trials were ethanol, dimethyl-sulphoxide (DMSO), dimethyl-formamide (DMF), 2-methyl-2,4-pentanediol (MPD). Two other solvents used were tri-fluoroethanol (TFE) and methanol, both of which have been shown from CD experiments to induce α-helix. All of the solvents were of (AR) grade. A stock solution of sCT (50mg/ml) was diluted with an equal volume of the organic solvent within the boiling point tube to produce an approximately 50% (vol/vol) solution before being heated as described in the experimental set up. The final protein concentration was 25 mg/ml. In another set of experiments all the organic solvents, with one exception, were added directly to lyophilised sCT to produce an approximately 100% solution. The exception to this was methanol

which has a boiling point of 64.7°C and was not used. Once again the final sCT concentration was 25mg/ml. An additional set of experiments were carried out to study the effect of pH on sCT's solubility in a range of ethanol concentrations from 10% to 50% under these experimental conditions.

4.4. Crystallisation by slow evaporation

This is a method which has often been applied in small molecule crystallisation to produce supersaturated solutions. Once again this method was applied only to the native hormone (sCT) and not to the two derivatives. The polypeptide was shown in small scale trials to be soluble in a variety of organic solvents. A set of experiments were set up using the solvents, ethanol, methanol, DMSO, DMF and trifluoroethanol. A 1ml solution of sCT in each of the 5 solvents was placed in a small glass test tube. The concentration of sCT was 2mg/ml. The top of the test tubes were sealed with parafilm. A small puncture in the seal was introduced using a Hamilton syringe to allow the solvents to evaporate slowly. The test tubes were kept at room temperature (23°C) and inspected regularly over a period of one month or until the solvent had almost completely evaporated.

4.5.Crystallisation by hanging drop

This method was applied to the attempted crystallisation of sCT and it's two derivatives and also to the co-crystallisation of equine PlA2 with an inhibitor molecule.

Experimental set up

The crystallisation boxes used were Linbro (Flow laboratories, McLean VA, USA.) 24-well tissue culture plates. The reservoirs contained 1 ml

of the precipitant solution. The drops were placed on siliconised coverslips (1cm x 1cm). The drop volume was 10 μl and the drops were formed by mixing 5μl of the protein solution with 5μl of the reservoir solution. The rims of the wells were lined with a silicone grease and the coverslip inverted over the reservoir. The concentration of calcitonin in the drop was 10 mg/ml. The buffers used were 100 mM in concentration and the pH range for their use is given in Table 4.1. The experiments were all performed and stored at room temperature, approximately 23°C. The plates were inspected initially after 4 days and subsequently every 7 days. The plates were kept for up to 4 weeks in order to ensure equilibrium had been reached. The plates were examined using a light microscope. For the more viscous precipitants, such as the PEG's and MPD the reservoir solutions were made up in Eppendorf tubes which were mixed thoroughly then centrifuged for 5 minutes at 13 000 rpm on a bench centrifuge to remove debris.

The methodology used in the attempted crystallisation of calcitonin and it's derivatives was initially to perform broad screen experiments where the precipitant concentration varied in relatively large steps. This was carried out over a broad pH range. The initial conditions of the trials were determined by small scale trials. These were performed by the direct addition of precipitant using a Hamilton syringe while viewing the drop under a microscope for precipitation. Results from the initial broad screening provided well defined boundaries for precipitation as a function of precipitant concentration and pH. Further crystallisation experiments involved expanding conditions between those yielding clear droplets and precipitate. Finally the effects of additives, such as small amounts of organic solvents and detergents added to the crystallisation droplets, was studied.

4.6 Histochemical staining of sCT aggregates with Congo Red.

The peptide precipitates obtained from vapour diffusion experiments were gently washed from the siliconised coverslips onto a depression slide using the well solution against which the drop had been equilibrated. Excess solution was carefully removed using a Hamilton syringe while viewing under a dissection microscope. Congo Red staining was carried out as described by the method of Puchtler, Sweat and Levine (1962). The solution around the aggregates was replaced with 80% ethanol saturated with NaCl for 20 minutes. This solution was next replaced by the stock stain solution (80% ethanol saturated with Congo Red and sodium chloride) and incubated for a further 20 minutes. Excess stain was removed and the aggregates washed three times in absolute alcohol. The stained aggregates were subsequently viewed under a microscope using cross polarised light.

4.7 Transmission electron microscopy of sCT aggregates

A single drop of a 10mg/ml solution of sCT in water was placed on carbon coated, low viscosity nitrocellulose treated copper grids. This was allowed to dry at room temperature before being negative stained with 2% (w/v) aqueous uranyl acetate. The peptide assemblies were observed in a Philips EM 400 T. E. M.. This same protocol was carried out for a 10mg/ml solution of melittin, an α-helical peptide from bee venom (Terwilliger et al., 1982).

4.8 Scanning electron microsocopy of sCT aggregates

Peptide micelles obtained from crystallisation experiments of sCT were examined using SEM. These micelles were removed from the siliconised coverslips by rinsing with 0.1 M sodium cacodylate-HCl

buffer (pH 7.3) onto a depression slide while viewing under a dissecting microscope. The peptide micelles were fixed with 3% gluteraldehyde in 0.1 M sodium cacodylate-HCl buffer (pH 7.3). The samples were then washed three times with excess buffer before being transferred onto a millipore filter (pore size 2μM). The filter paper envelope was then dehydrated through a graded acetone series of 50%, 70%, 90%, 100%. The sample was next critical point dried as described by the method of Cohen (1979) and then sputter coated with a 20nm film of a 60/40 gold/palladium mixture as described by the method of Echlin (1975). The specimens were subsequently viewed in a Philips 505 scanning electron microscope. A similar procedure was applied to amorphous precipitate specimens. Hanging drops containing amorphous precipitate of sCT were rinsed directly onto the filter paper with buffer. The specimen was then dehydrated and coated as described above before being viewed under the scanning electron microscope.

4.9 Co-crystallisation of an inhibitor with equine PLA-2

The equine PLA-2 and the inhibitor compounds were generous gifts of Sandoz AG. All were obtained in the form of lyophilised powder. The hanging drop method as described previously was used in the co-crystallisation. Six different compounds, which had inhibitory activity towards PLA-2, were made up as 1.3 M stock solutions in DMSO. The protein stock solution contained PLA-2 (20mg/ml), 30% ammonium sulphate, 50 mM imidazole-HCl pH 7, 5mM CaCl2. Using a Hamilton syringe, 1μl of the stock drug solution was added to 60.5 μl of the protein solution, resulting in a 20mM solution of the drug. In the control experiment 1μl of DMSO alone was added to the protein solution. The protein-drug mixture was whirlimixed and allowed to

stand for 2 hours at room temperature. The solution was then spun on a bench centrifuge for 2 minutes at 13 000 rpm to clarify the solution. Before plating out, 3.5μl of a seeding solution was added to the protein drug mixtures using a Hamilton syringe and the Eppendorfs gently agitated. The seeding solution was prepared by crushing several large crystals of equine PLA-2 in 1ml of their mother liquor, 50mM Imidazole-HCl pH 7, 2mM $CaCl_2$ and 30% ammonium sulphate. The plates were prepared as described previously using an automated pipetting station. The initial trials were performed at pH 7 and the wells contained 29, 30, 31, 32, 33 and 34% ammonium sulphate. The drop volume was 10μl of the protein-drug and seed mixture and the plates were stored at room temperature. Only the conditions with one drug, shown in figure 4.2, produced crystals and were expanded. A range of pH's from 6.8 to 7.2 were studied.

4.10 Results

4.10.1 Purity of salmon calcitonin

The results of the amino acid analysis are shown in table 4.2. The amino acid composition is identical to that expected for salmon calcitonin indicating that the peptide is indeed absolutely pure and chemically identical to salmon calcitonin.

4.10.2 Slow cooling experiments

Effect of calcitonin concentration

With the exception of pH 7, calcitonin was soluble at 70°C up to a

Phospholipase Inhibitor,
Molecular Formula $C_{30}H_{41}BrO_4$

Molecular Weight: 545.6
Racemic
Soluble in DMSO.

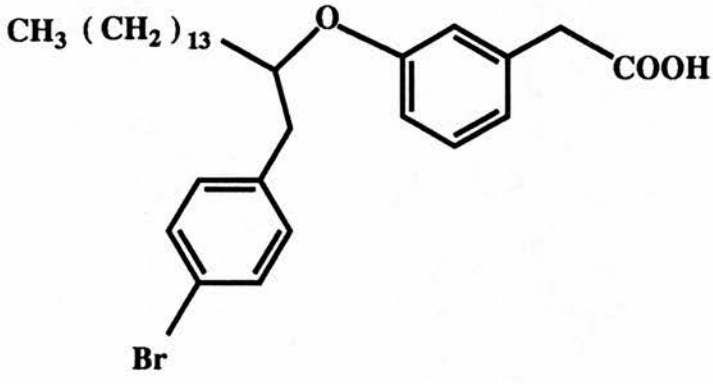


Figure 4.2. The chemical structure of the PLA-2 inhibitory compound used in the co-crystallisation with equine PLA-2.

Table 4.2 Amino Acid Analysis of Salmon Calcitonin. Hydrolysis with 6 N HCl for 24 hours.

Amino Acid	Molar ammount	Expected
Alanine	0	0
Arginine	1.045	1
Asx	2.186	2
Cysteine	2.027	2
Glx	3.028	3
Glycine	3.16	3
Histidine	1.244	1
Isoleucine	0	0
Leucine	5.307	5
Lysine	2.134	2
Methionine	0.234	0
Phenylalanine	0	0
Proline	1.985	2
Serine	3.903	4
Threonine	4.858	5
Tyrosine	1.021	1

LKB 4400 with Norleucine internal standard.

concentration of 100 mg/ml at all pH's. It was observed that calcitonin dissolved most easily in the more acidic buffers. At pH 7.0 a small fraction of calcitonin did not dissolve and was removed by spinning. Thus the values of all subsequent dilutions at pH 7 for this experiment were overestimated. This was not observed at pH 8.0. On cooling, all the 100 and 50 mg/ml solutions had precipitated at all pH's. However the nature of the precipitate was significantly dependent upon pH. All the acidic solutions had formed precipitates that consisted of small spheres (diameter approximately 0.1mm). These spheres were soft on crushing and were similar to those obtained later with the hanging drop experiments as shown later in Plate 4.1 and are hereafter called peptide micelles. The precipitates at pH 7 and 8 were floccular in appearance and had sunk to the bottom of the boiling point tubes. There was no precipitation at 20 mg/ml or below at all pH's. All subsequent experiments were carried out with calcitonin concentrations of 25 mg/ml.

Effect of divalent metal ion

No precipitation was detected at any molar ratio of either Ca^{2+} or Zn^{2+} at a peptide concentration of 25 mg/ml. There was therefore no detectable pH effect. No further experiments were carried out with divalent metal ions at higher calcitonin concentrations or in the presence of other precipitants.

Effect of organic solvents

The results of the effects of the different organic solvents on the solubilty of sCT are summarised in Table 4.3. When the peptide was

Table 4.3
Solubility of sCT in organic solvents using the slow cooling method for crystallisation. The method used was as described in the methods section. The concentration of sCT in these solvents was 25 mg/ml.

Solvent	% (vol/vol)	Observation
Ethanol	50	gels & peptide micelles
	100	clear
Methanol	50	gel
DMSO	50	gel
	100	clear
DMF	50	gel
	100	clear
MPD	50	clear
	100	clear
TFE	50	clear
	100	clear

dissolved in neat solvent no precipitates were obtained after cooling. However, with the exception of MPD and TFE, 50% aqueous solutions of the solvents produced gelation. Peptide micelles were observed in 50% ethanol solutions. The effect of pH on the solubility of sCT in ethanol solutions are summarised in table 4.4. There was no form of precipitation in the absence of ethanol and at the more acidic pH's (pH 4 and 5) in the presence of 10% ethanol. However, at pH 6 the solution had a viscous opaque gel-like appearance and at pH 7 and 8 the solutions contained spherical aggregates that had a familiar gel-like appearance. These were found at all pH's at only 50% ethanol.

4.10.3 Slow evaporation experiments

The solvents ethanol, methanol and trifluoroethanol evaporated completely after about 10 days leaving the sCT as an oily residue at the base of the test tube. The polypeptide did not appear to come out of solution until these solvents had almost completely evaporated. The other two solvents DMSO and DMF had evaporated to about half the original volume without any signs of precipitation in the period of one month.

4.10.4 Hanging drop experiments

1. Salts

The solubility profiles of sCT in the hanging drop method in the presence of various salts as the precipitating agent are shown in Figures 4.3a-d. They also indicate the different forms of precipitate obtained. Two forms of precipitate are illustrated in plates 4.1a-b. The results show a solubility minimum in the pH range between pH 6 - 8 the exact minimum being dependent on the salt used. Amorphous

Table 4.4
Effect of pH on solubility of sCT in ethanol solutions. The slow cooling experiment was carried out as described in the methods section. The peptide concentration was 25 mg/ml.

Ethanol concentration (%)	pH	Observation
0	4	clear
	5	clear
	6	clear
	7	clear
	8	clear
10	4	clear
	5	clear
	6	opaque gel
	7	opaque gel
	8	opaque gel
20	4	clear
	5	clear
	6	opaque gel
	7	opaque gel
	8	opaque gel
30	4	clear
	5	clear
	6	opaque gel
	7	opaque gel
	8	opaque gel
40	4	clear
	5	clear
	6	opaque gel
	7	opaque gel
	8	opaque gel
50	4	opaque gel
	5	opaque gel
	6	opaque gel
	7	opaque gel
	8	opaque gel

Figure 4.3 a-d. pH dependent solubility profiles of sCT with a variety of salts as precipitant. Vapour diffusion experiments were carried out as described in the methods section. The peptide concentration in the drop was 10mg/ml. The salts used were as follows: A ammonium sulphate; B-zinc acetate; C-lithium chloride; D-sodium chloride.

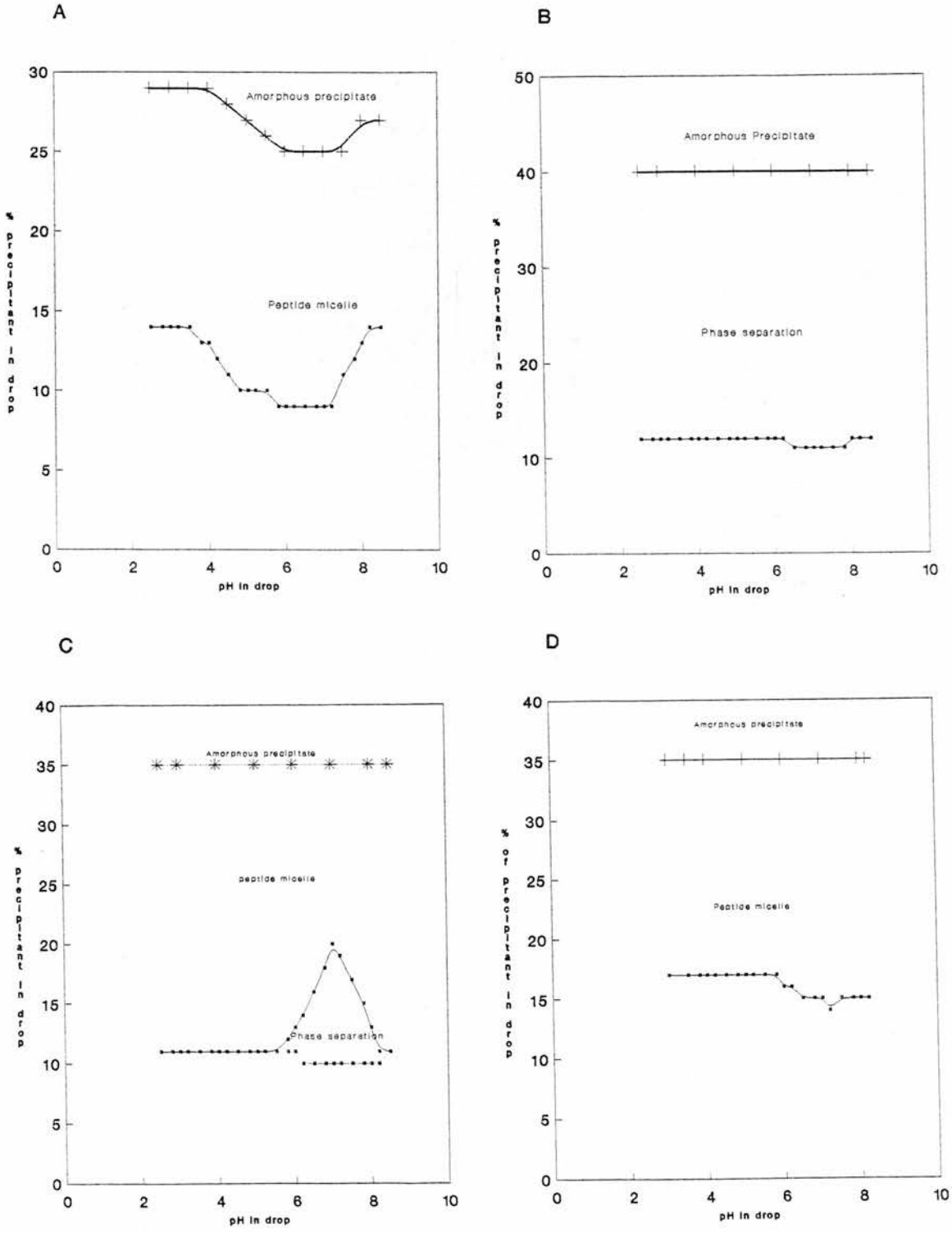


Plate 4.1a.
Peptide micelles formed by sCT during vapour diffusion experiments. Magnification x 25.

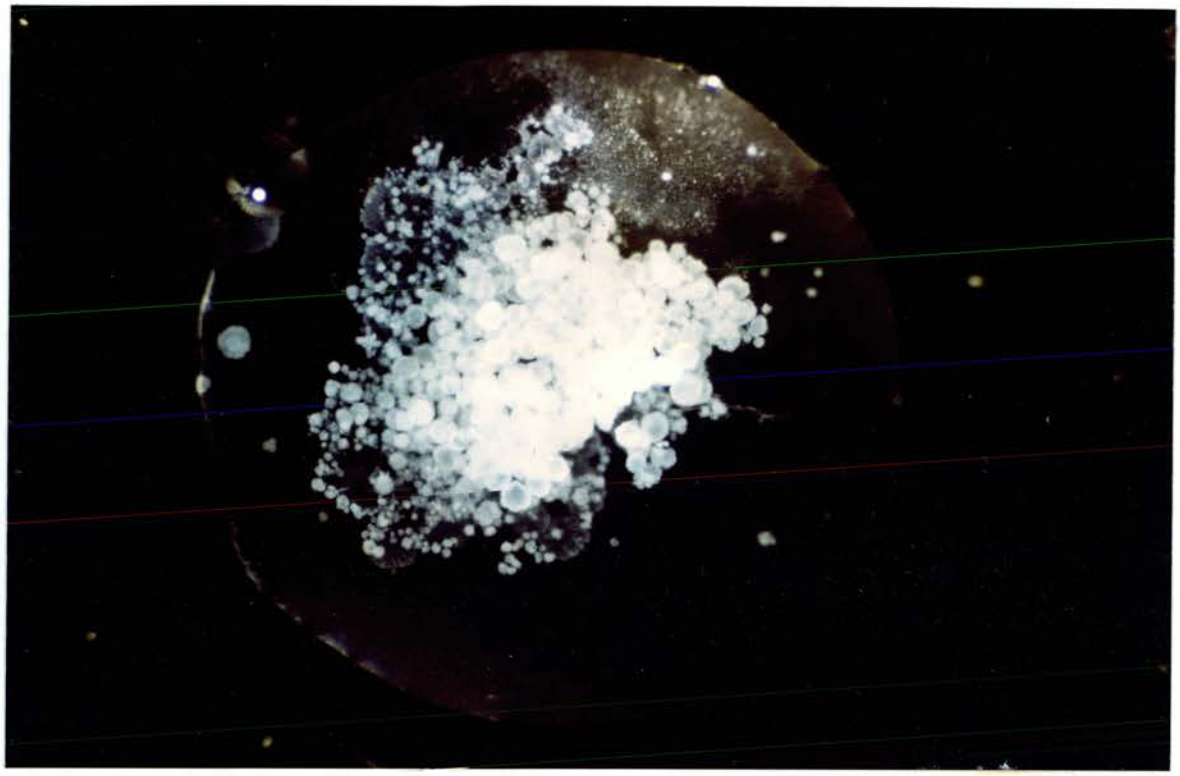
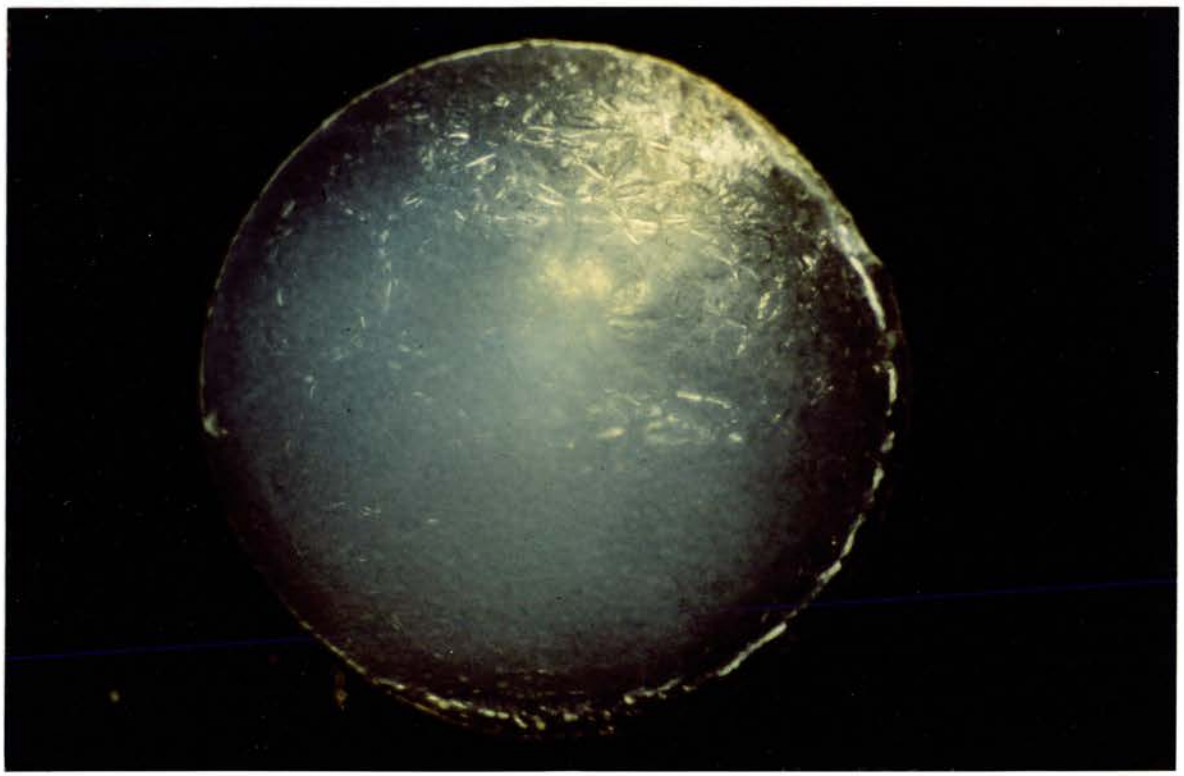


Plate 4.1b
Surface skins formed by sCT during vapour diffusion experiments. Magnification x 25.



precipitates are found in the presence of high levels of all the precipitants tested. Peptide micelles are obtained at lower levels of precipitant in three of the four salts investigated; namely ammonium sulphate, lithium chloride and sodium chloride. In two salts a phase separation was obtained at low levels of precipitant; namely zinc acetate at all pH's and lithium chloride between pH 6.2 - 8.2. At no point along the solubility minima of all the precipitants were crystalline precipitates obtained.

2. PEG's

The solubility profiles of sCT in the hanging drop method with various polyethylene glycols as the precipitating agent are shown in Figures 4.4 a-d. A solubility minimum is observed in all cases around pH 7. Only amorphous precipitate was observed with this type of precipitant. No peptide micelles, surface skins or phase separations were observed, even at the solubility minima. There was no general correlation between solubility of sCT and the average molecular size of the PEG used.

3. 2-Methyl-2,4-pentanediol

The solubility profile of sCT in the hanging drop method with MPD as the precipitating agent is shown in Figure 4.5. The peptide appears to be particularily soluble in the presence of high levels of MPD. Only amorphous precipitate was obtained and only a slight pH effect was observed.

4. Additives

The effects of additives on the solubility profile of sCT with a

Figure 4.4 a-d. pH dependent solubility profiles of sCT with a variety of different polyethylene glycols as precipitant. Vapour diffusion experiments were carried out as described in the methods section. The peptide concentration in the drop was 10mg/ml. The PEG's used were as follows: A - PEG 1 550; B - PEG 4 000; C - PEG 10 000; PEG 20 000. The numbers refer to the average molecular weight of the PEG.

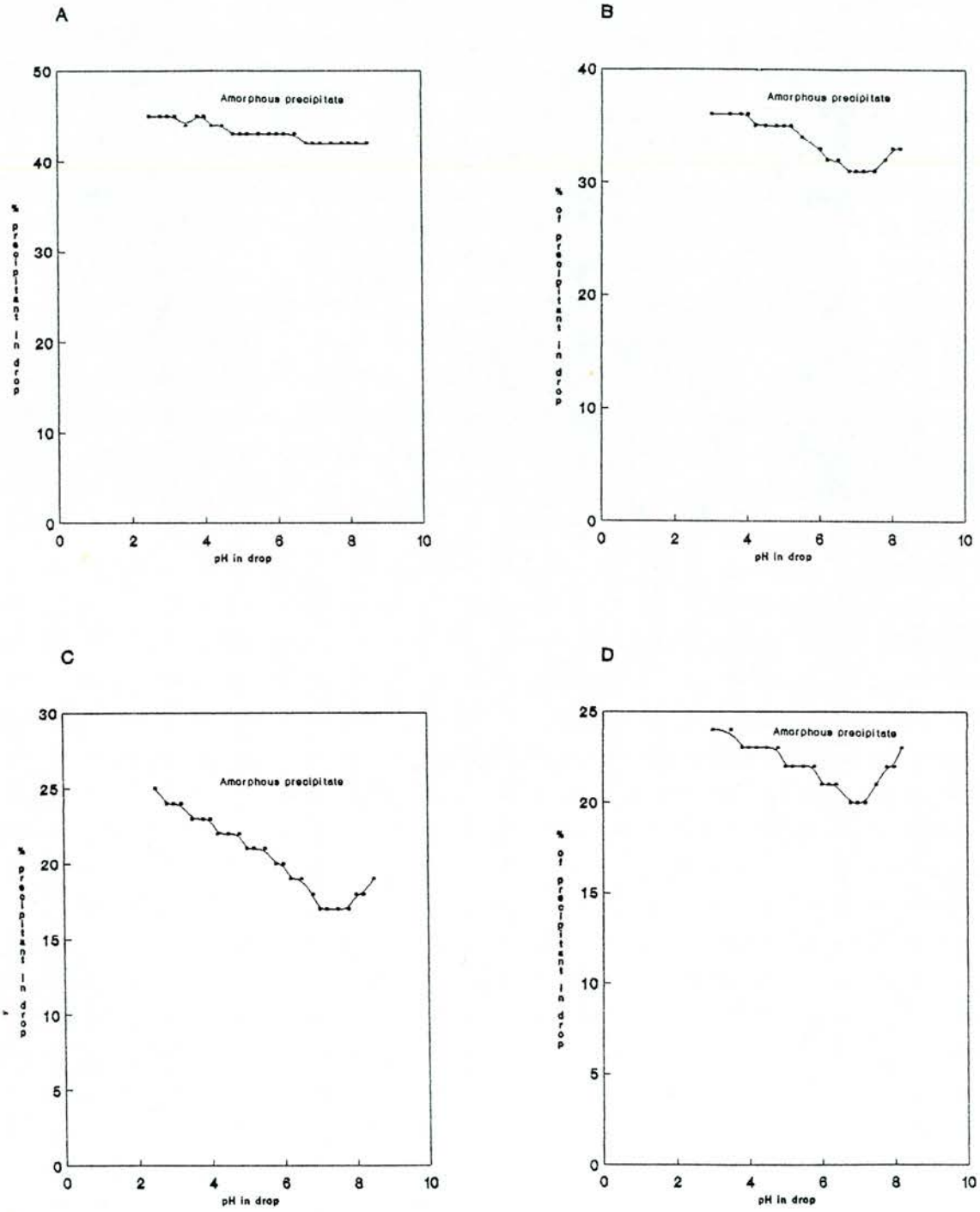
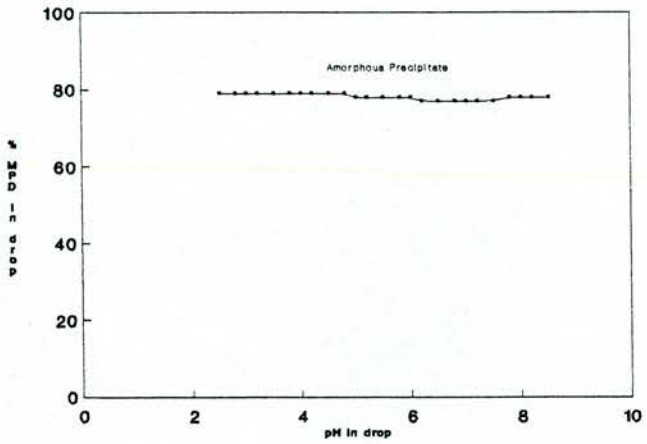


Figure 4.5 pH dependent solubility profile of sCT with methylpentane-diol (MPD) as precipitant. Vapour diffusion experiments were carried out as described in the methods section. The peptide concentration in the drop was 10mg/ml.



variety of precipitants at pH 7.2 are shown in Figure 4.6 a-d. The two detergents used (octyl-glucoside and SDS) produced opposing results. Firstly the effect of increasing octyl-glucoside concentration, Figure 4.6a, resulted in a decreased solubility of sCT in PEG 4000. However, the presence of SDS, even as low as 0.25 % (w/v), increased solubility of sCT in PEG 4000 at pH 7.2. There was no observed effect of SDS on the nature of the precipitant at supersaturation. The presence of octyl-gucoside also promoted the formation of a skin on the surface of the drop as indicated in Figure 4.6a and shown in plate 4.1b. The effect of the phospholipid, DMPG, on the solubility of sCT in PEG 4000 was to alter not only the solubility but the form of the precipitate obtained as shown in figure 4.6 d. The peptide appeared less soluble at all pH's and also promoted the formation of peptide micelles as shown in plate 4.1. This is in marked contrast to the effect of the other α-helix promoting substance SDS, which increased sCT solubility at all pH's. DMSO appeared to have no effect on the solubilty of sCT at the concentrations used in the experiment as shown in figure 4.6 c.

5.Derivatives

The solubility profiles of the two derivatives with a variety of precipitating agents in the hanging drop method are shown in Figures 4.7 a-f. No peptide micelles or surface skins were observed with either of these derivatives under any of the conditions tested. The two derivatives had very similar solubilty profiles in the presence of a variety of precipitants. However both peptides appeared more soluble and resistant to precipitation than sCT in the presence of the three precipitants tested. However, like sCT both the derivatives appeared to possess a slight solubility minimum around pH 7.

Figure 4.6 a-d. Solubility profiles of sCT in PEG 4000 at pH 7.2 showing the concentration dependent effects of a number of additives. Vapour diffusion experiments were carried out as described in the methods section. The peptide concentration in the drop was 10mg/ml. The additives used in these studies were as follows: A - Octyl-glucoside; B - SDS: C - DMSO: D - DMPG.

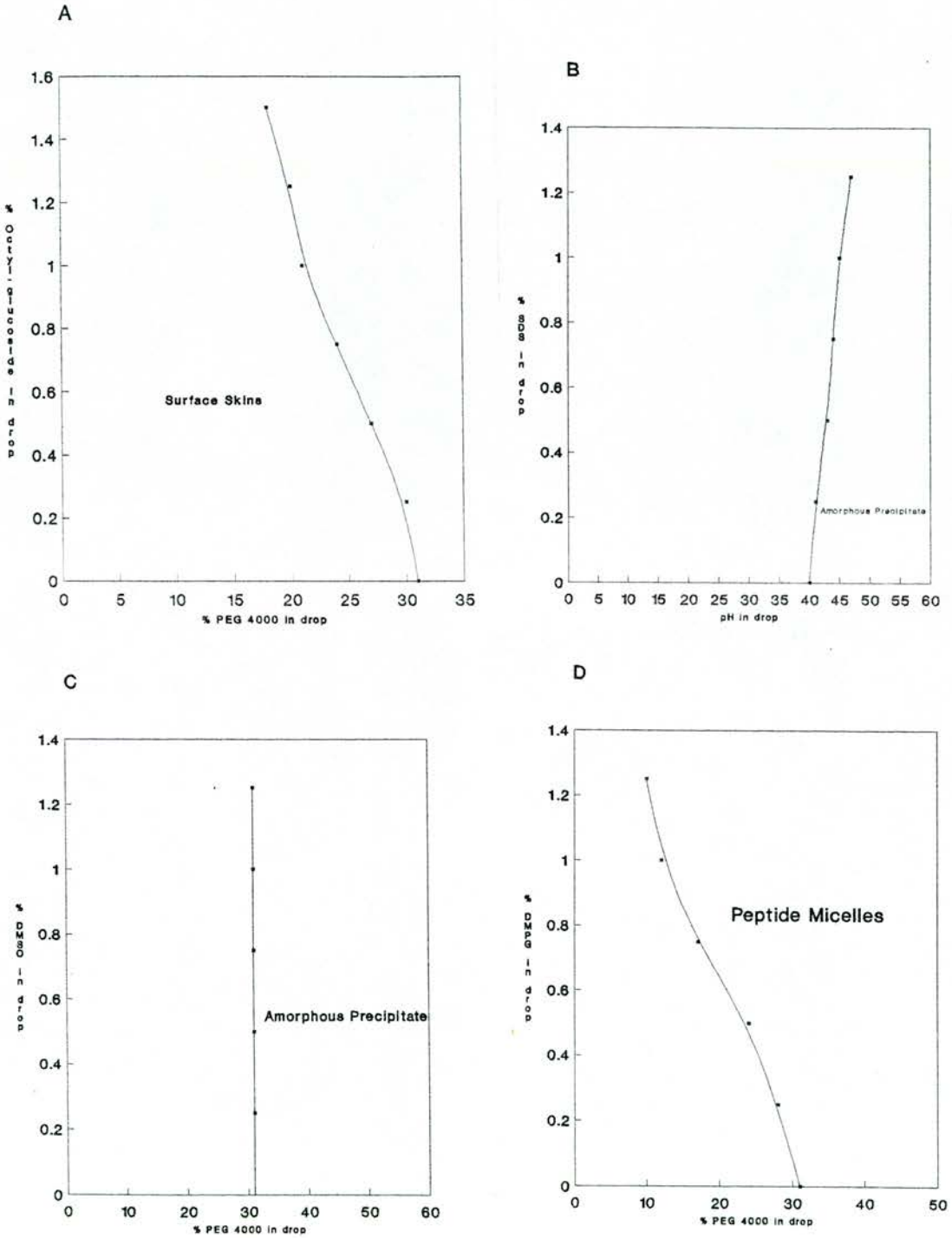


Figure 4.7 a-c. pH dependent solubility profiles of derivative 1 with a variety of precipitants. Vapour diffusion experiments were carried out as described in the methods section. The peptide concentration in the drop was 10mg/ml. The precipitants used were as follows: A-ammonium sulphate; B-PEG 20 000 C-MPD.

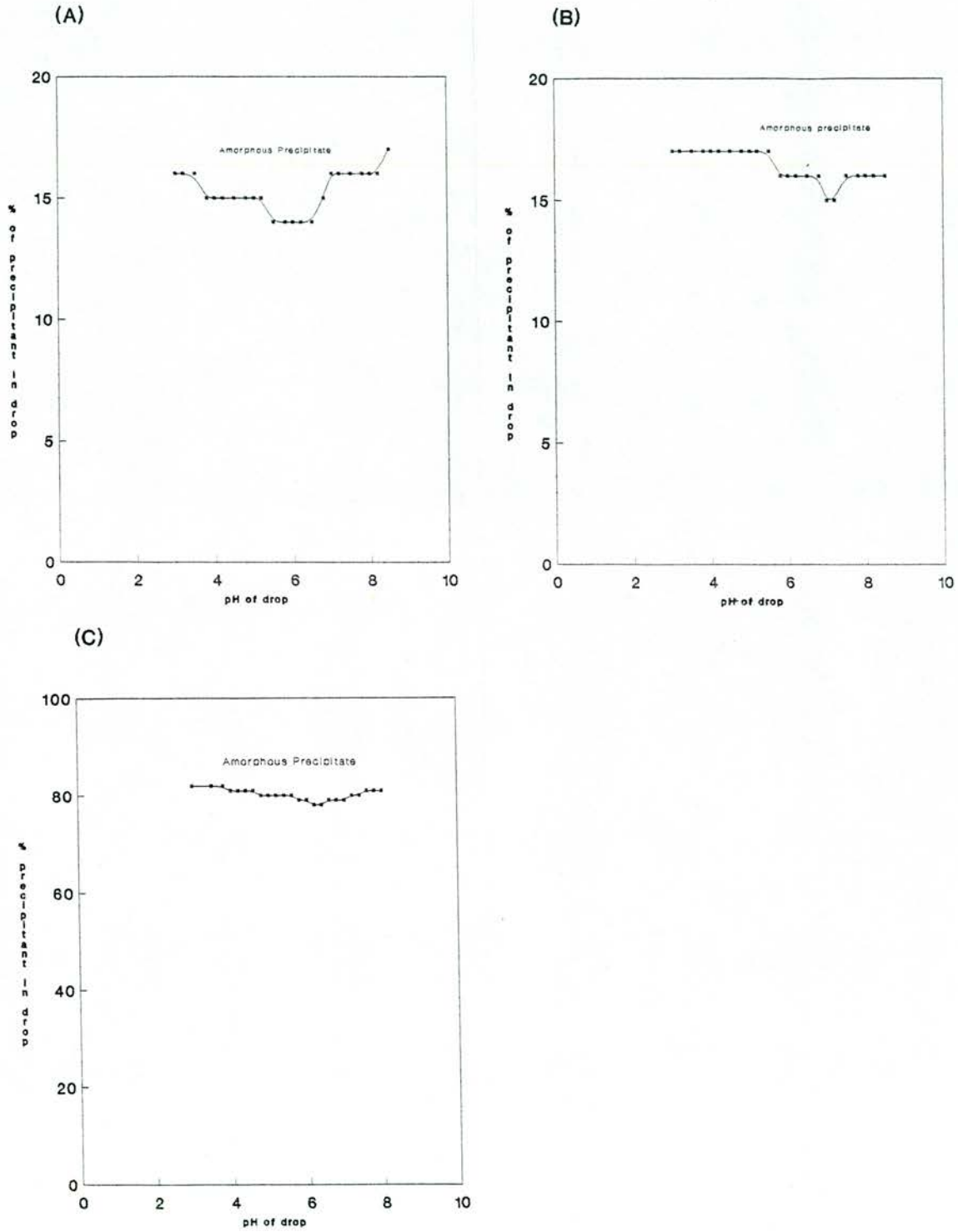
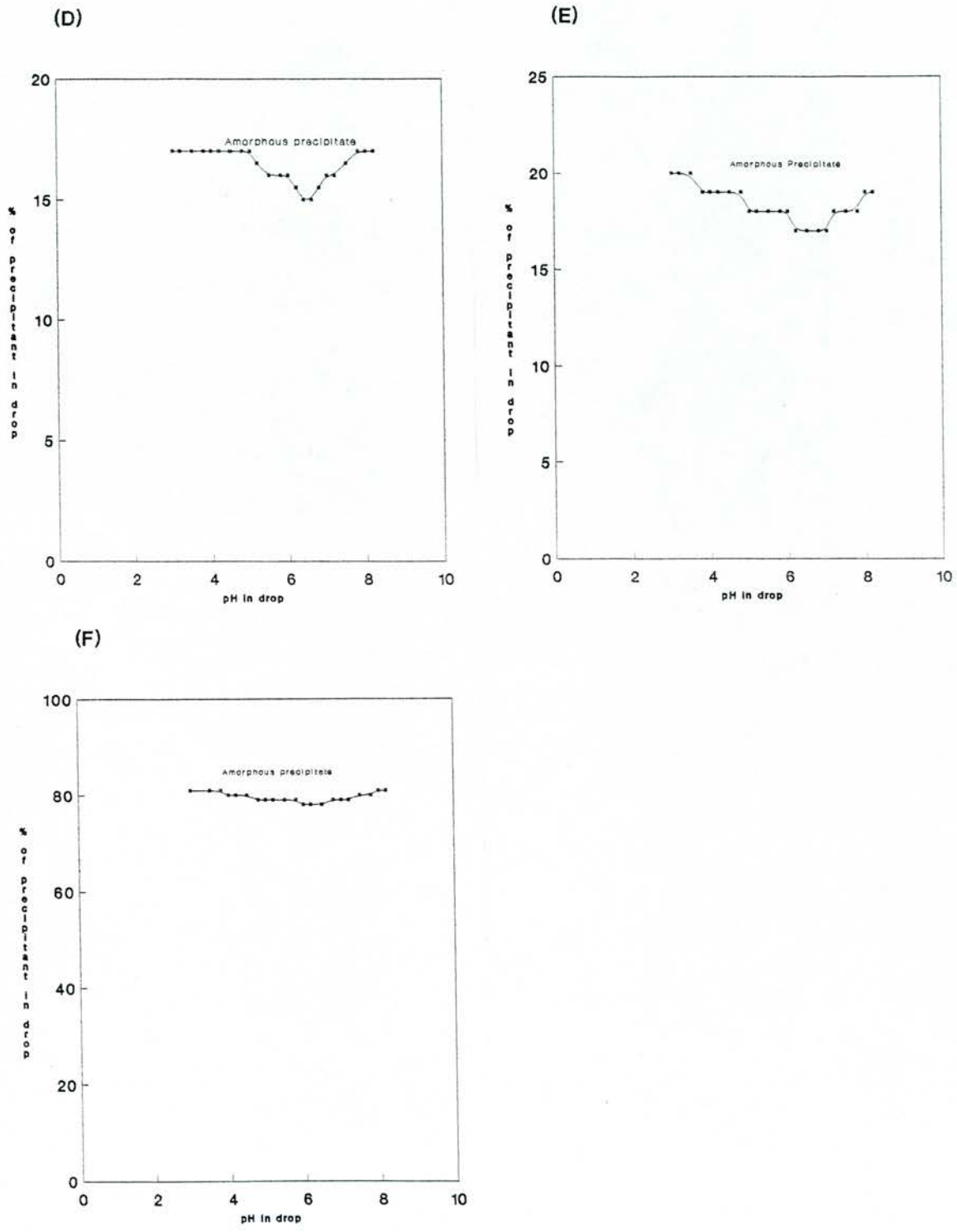


Figure 4.7 d-f. pH dependent solubility profiles of derivative 2 with a variety of precipitants. Vapour diffusion experiments were carried out as described in the methods section. The peptide concentration in the drop was 10mg/ml. The precipitants used were as follows: D-ammonium sulphate; E-PEG 20 000; F-MPD.



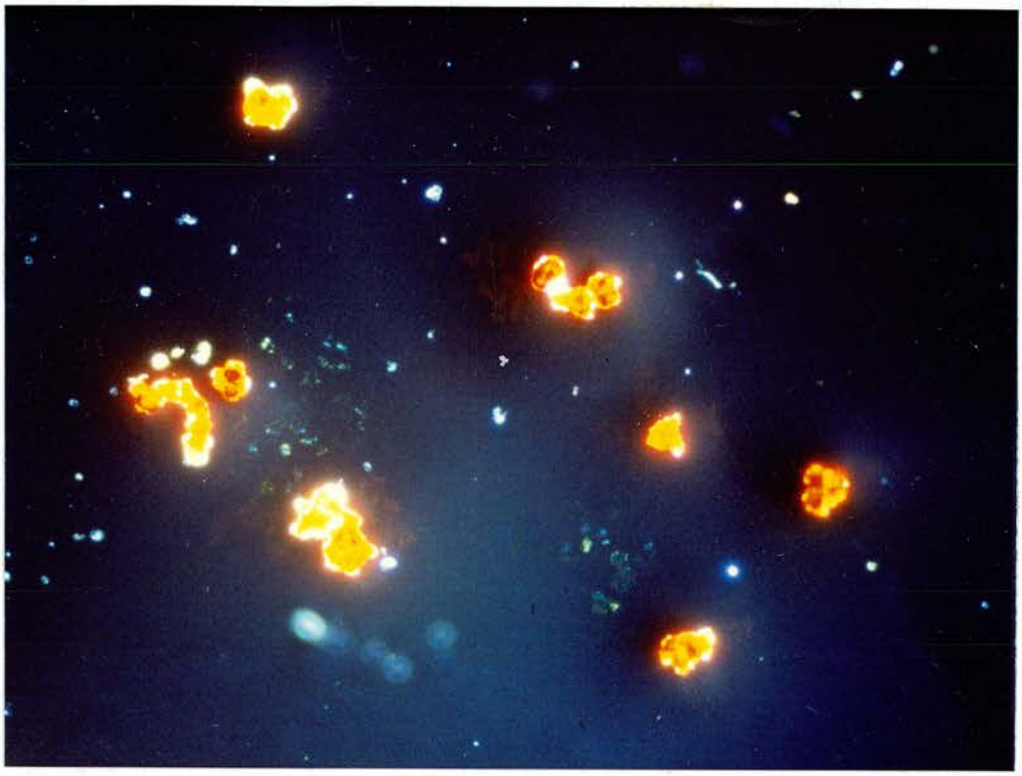
4.10.5 Histochemical staining of sCT aggregates

Due to the fragile nature of both amorphous precipitate and the surface skins it was not possible to determine their ability to stain with Congo red. Peptide micelles were shown to retain stain after washing, these stained aggregates also showed a green birefringence under cross polarised light as shown in plate 4.2 a-b. Despite the slight differences of morphology, peptide micelles obtained at a variety of pH's and with a variety of precipitants were tested and shown to stain positively. Peptide micelles produced with a variety of precipitants as indicated in figs 4.3 - 4.6 were tested and shown to stain a red-pink colour as shown in plate 4.2 a. When these stained aggregates were crushed and viewed at higher power a fibrous structure was observed as shown in plate 4.3. This is indicative of amyloid formation.

4.10.6 Transmission electron microscopy

Fibrils of approximately 60 Å in diameter and of maximum length 20000 Å were observed after negative staining as shown in plate 4.4. These are of similar dimensions to the fibrils of ß-protein fragments (Halverson *et al.*, 1990; Kirschner *et al.*, 1987; Burdick *et al.*, 1992). These fibrils were single and unbranching but were occasionally observed to associate along their long axis in ribbon type structures as shown in plate 4.5 similar to those of ß-protein peptides (Fraser *et al.*, 1991). There was evidence of twisted fibrils as shown in plate 4.6. No fibrils were obtained for the control peptide, mellittin as shown in plate 4.7.

Plate 4.2
a) Peptide micelles after staining with Congo red under normal illumination. Micelles were shown to stain a deep red-pink colour. Magnification x315. b) Peptide micelles after staining under cross-polarised illumination. A green-blue birefringence is observed typical of amyloid. Magnification x375.



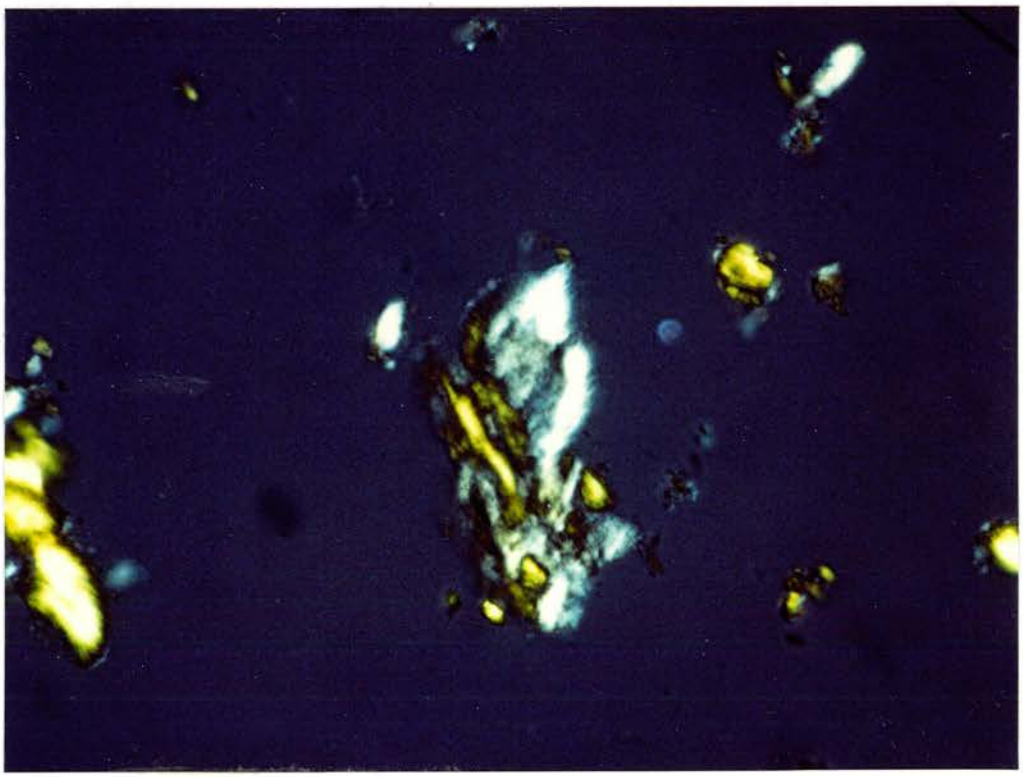


Plate 4.3 Congo red stained peptide micelle under cross polarised illumination. The fibrous sub-structure of the micelle is apparent. Magnification x375.

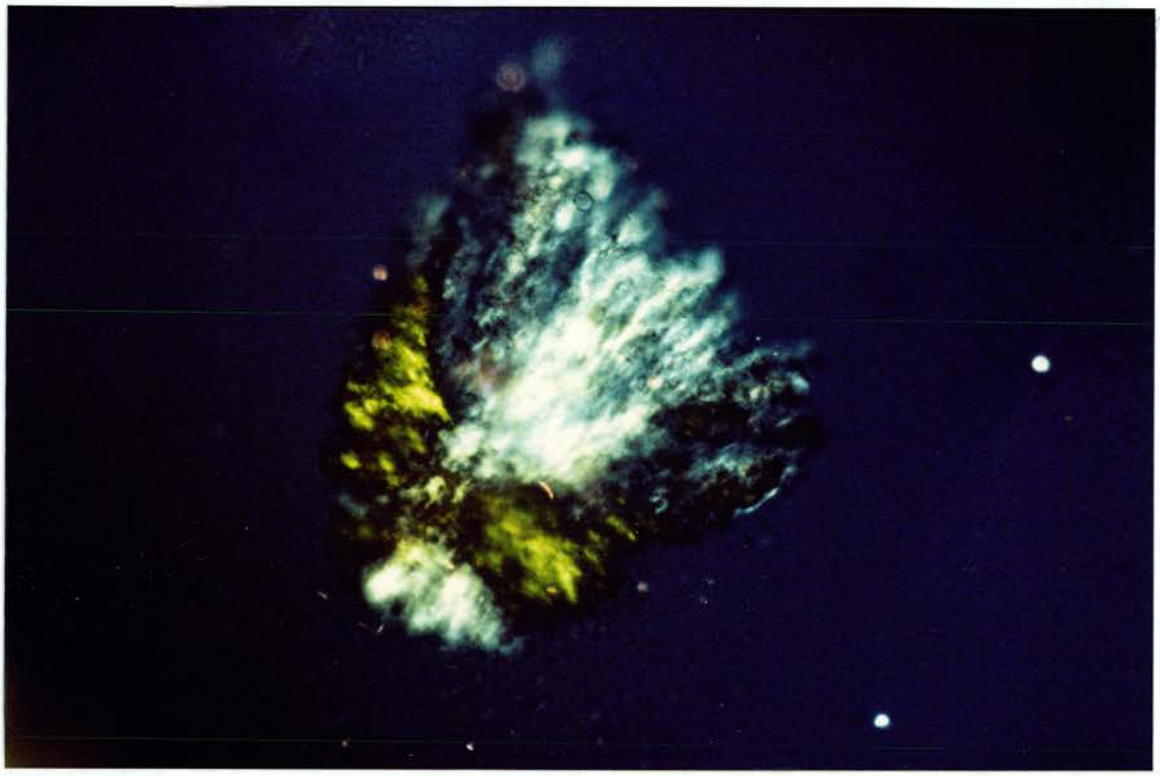


Plate 4.4
Negatively Stained electron-micrographs of sCT fibrils. Staining with 2% uranyl acetate was performed as described in the methods section. Magnification X 100 000. Fibril diameter approximately 6 nm and of length upto 2000nm. Scale bar represents 100nm.

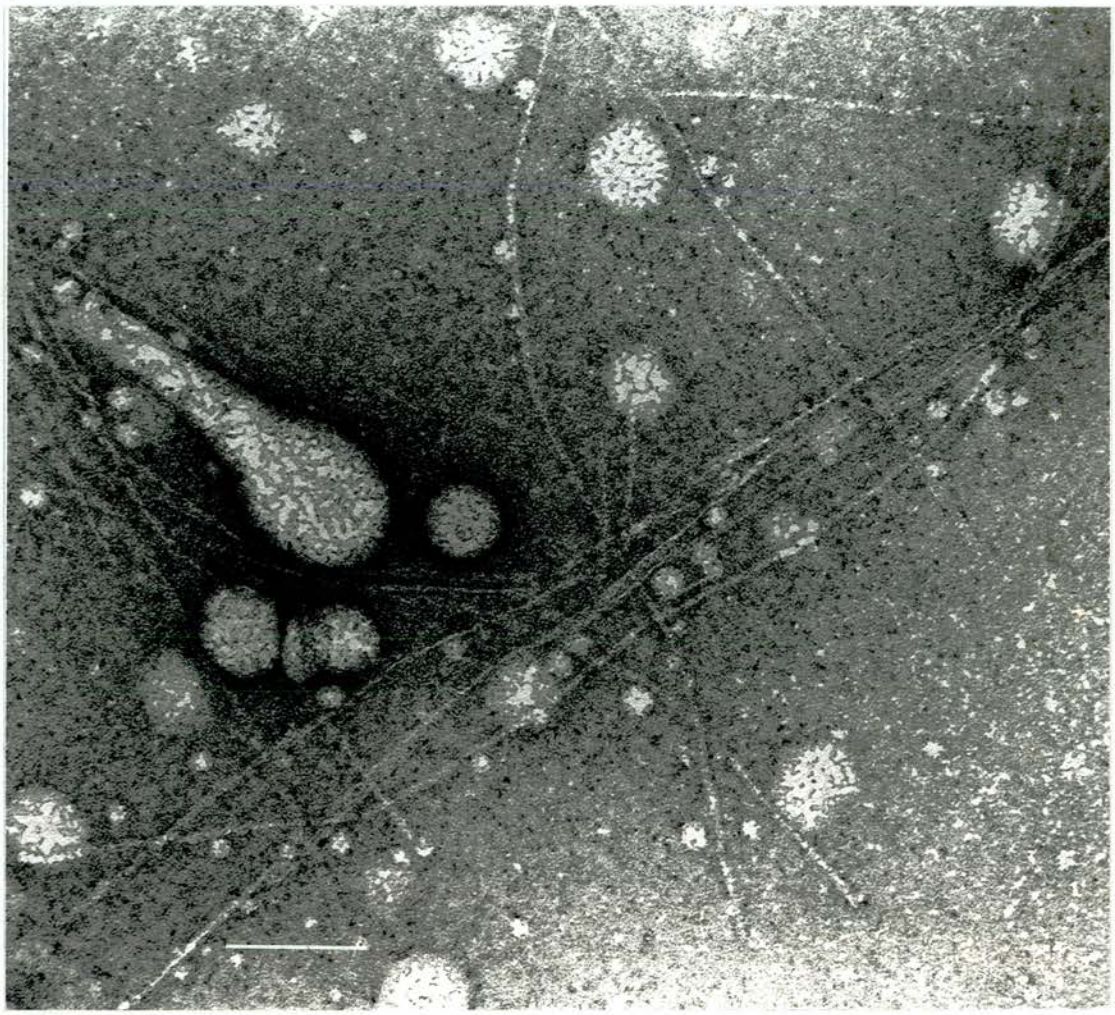


Plate 4.5
Negatively stained electron-micrographs of sCT fibrils illustrating the ribbon type aggregations of individual fibrils. Magnification x100 000, scale bar represents 100nm.

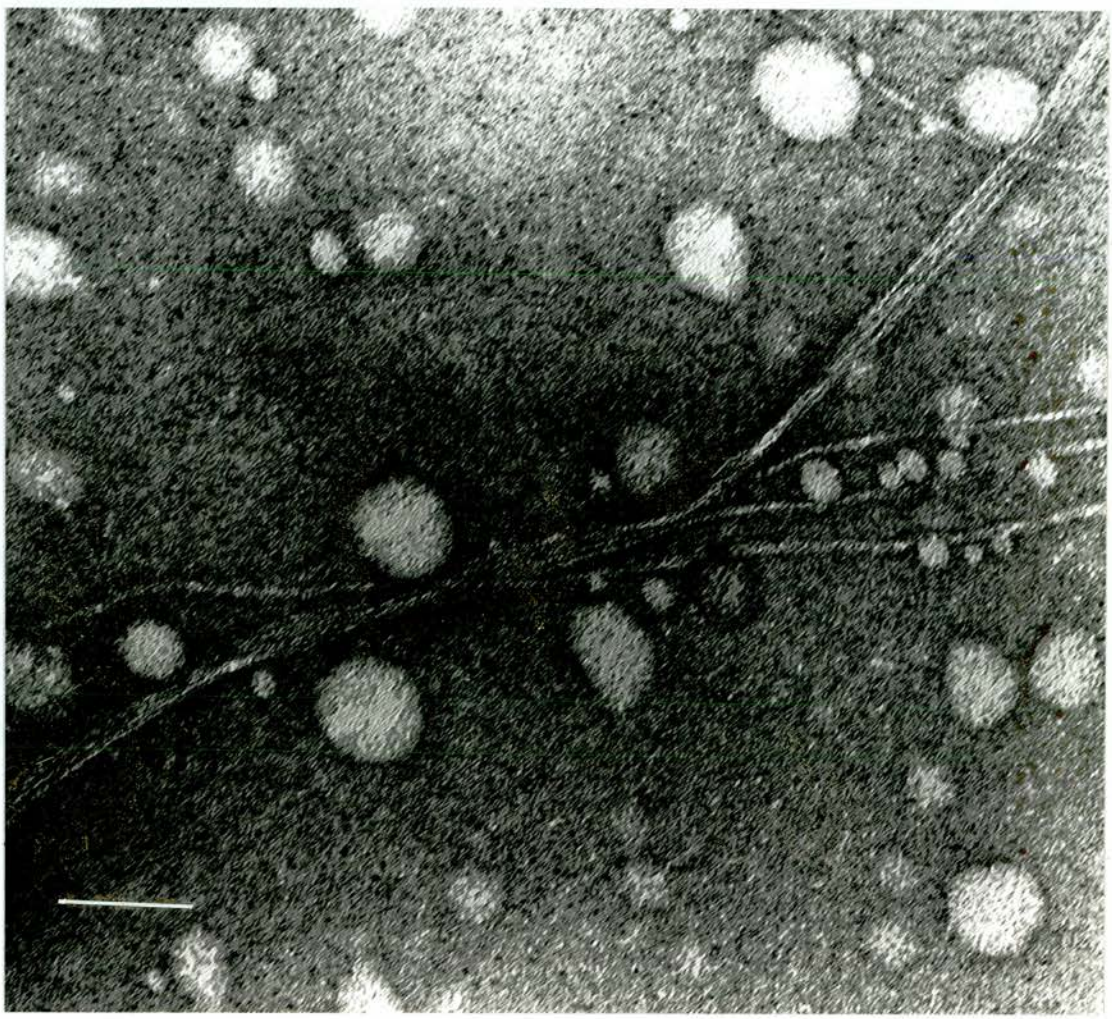


Plate 4.6
Negatively stained electron-micrograph of sCT fibrils illustrating the intertwining and twisted nature of pairs of fibrils.
Magnification x100 000, scale bar represents 100nm.

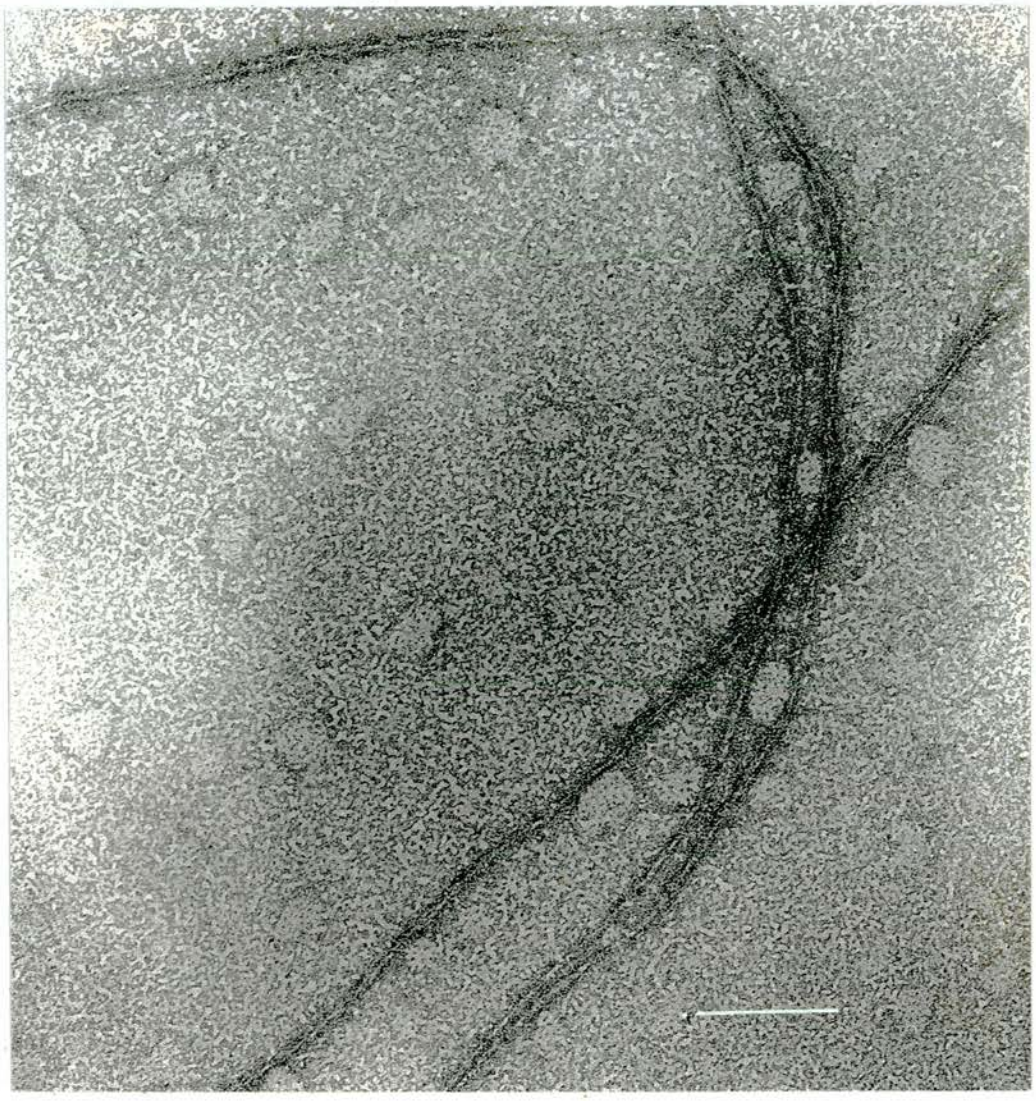
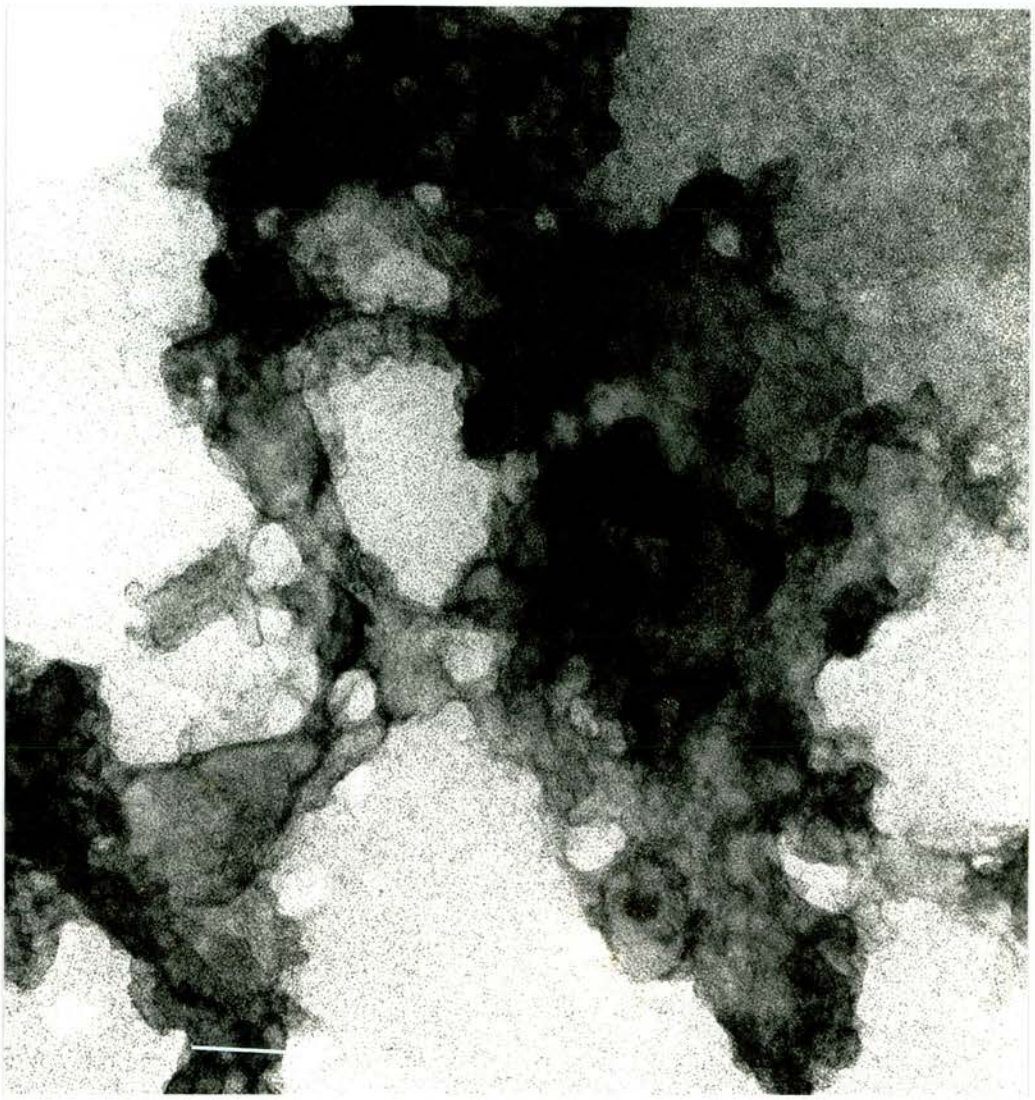


Plate 4.7
Negatively stained electron-micrograph of mellitin aggregates. The peptide (10mg/ml) was stained in exactly the same manner as sCT as described in the methods section. No fibrils were identifed. Magnification x188 000, scale bar represents 100nm.



4.10.7 Scanning electron microscopy

The peptide micelles produced in the vapour diffusion experiments are clearly fibrous in nature as shown in plate 4.8. The constituent fibrils appear to be on average approximately 150 nm in diameter as shown more clearly in plate 4.9. Thus these fibres must be composed of at least 25 of the 60 Å fibres which were observed in the T.E.M. across the diameter of the fibrils observed in the S.E.M..

Amorphous precipitate as shown in plate 4.10 is clearly not fibrous in morphology and does not form large aggregates.

4.10.8 Equine PLA-2 co-crystallisation experiments

In all of the experiments, except for the controls, it was clear that some fraction of all six of the hydrophobic PLA-2 inhibitors were dropping out of solution. This occurred even at the lowest ammonium sulphate concentration. The extent of the precipitate increased slightly with increasing salt until at higher salt concentrations, the protein or possibly protein with bound drug also dropped out of solution. In five cases at the higher salt concentrations only heavy precipitate was obtained that showed no signs of crystallinity. However in the case of the inhibitor shown in figure 4.2, a micro-crystalline precipitate was obtained between 32% and 34% salt which on expanding the conditions produced single crystals. The effect of DMSO on the control experiments was not to inhibit crystallisation. However the crystals were obtained at a slightly higher concentration, 32% ammonium sulphate, than the established conditions for the equine PLA-2. Large single crystals suitable for x-ray diffraction studies were obtained at pH 7.2 and 34% ammonium sulphate after 9 days. These were of a slightly different morphology to that of the native crystals. It

Plate 4.8
Scanning electron-micrographs of peptide micelles grown in vapour diffusion experiments at pH7.2, 30% (w/v) PEG 4000, 6% (w/v) DMPG. The micelles were prepared for microscopy as described in the methods section. Magnification x1295 scale bar represents 10μm.

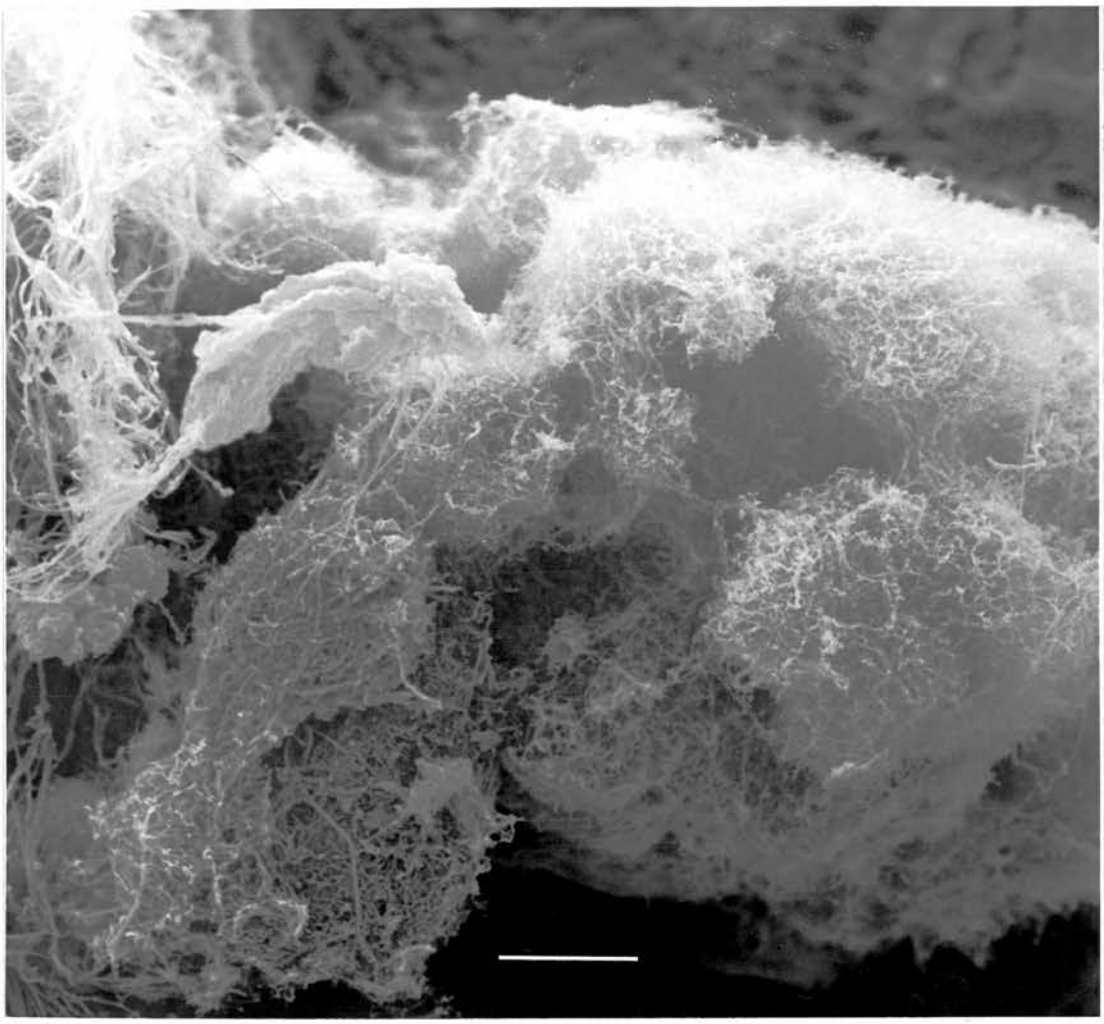


Plate 4.9
Scanning electron micrograph of peptide micelle fibrils. The micelles were grown in vapour diffusion crystallisation trials as previously indicated in legend 4.10. The micelles were prepared for microscopy as described in the methods section. Magnification x 10, 300, scale bar represents 1μm.

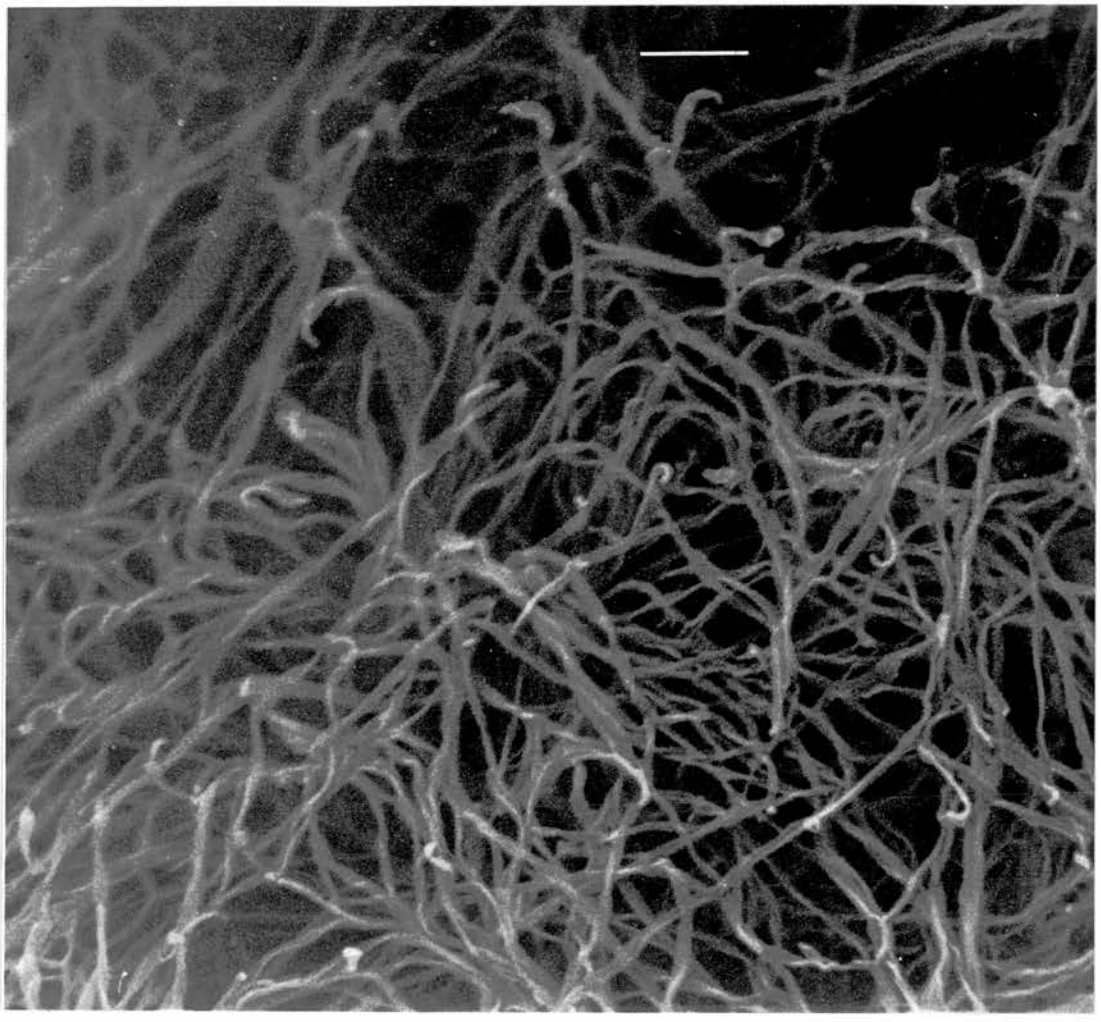


Plate 4.10
Scanning electron micrograph of amorphous precipitate. Amorphous precipitate of sCT was obtained during vapour diffusion experiments as described in the methods section. The precipitate was prepared for microscopy as described in the methods section. Magnification x1535, scale bar represents 10μm.

was hoped that this slight difference was due to the formation of the enzyme-inhibitor complex.

4.11 Discussion.

None of the methods applied to the crystallisation of sCT and two analogues has proved successful. These methods have previously proved successful in the crystallisation of a wide range of macromolecules; from whole virus particles (Hogle *et al.*, 1985) to membrane proteins (Kreusch *et al.*, 1991). The lack of success with salmon calcitonin cannot therefore be attributed to some flaw in the methods. Rather it is likely that there are some structural features or solution properties of sCT which have prevented crystallisation. Only non-crystalline material has been obtained which in the case of the peptide micelle has been shown to be far from amorphous. The polymorphic nature of these precipitates has made a systematic approach to the crystallisation of sCT difficult. These unusual precipitates have been produced both in the slow cooling experiments and in the vapour diffusion experiments further indicating that some consistent features of the solution properties are responsible for the formation of these unusual precipitates. There have been no previous reports of micellar aggregates with peptides. However, the formation of 'gels' has been observed in other amphipathic peptides at high concentrations (Epand *et al.*, 1989) and surface skins have been observed with peptide fragments of the A4/ß amyloid protein of Alzheimer's disease (Burdick *et al.*, 1992).

The nature of these unusual aggregations was examined by two independent techniques namely, histochemistry and electron microscopy, which have both confirmed the existence of amyloid. Amyloid deposits

are characterised by electron microscopy as consisting of fibres that are 4-10 nm in diameter (plate 4.4) and histologically as exhibiting green birefringence under polarised illumination when stained with congo red (plate 4.2b).

The formation of amyloid fibrils by calcitonin obviously has important implications for the interpretation of the crystallisation trials. It provides an explanation as to why calcitonin has resisted crystallisation. The crystallisation process requires that the monomer remain as an essentially monodisperse solution up to the point of superstaturation. Here it is obvious that fibrillogenesis is occurring before supersaturation since these fibrils can form even in the absence of precipitate such as during the preparation of the T.E.M. samples. Although the peptide can exist in solution as a monomer, at the high concentrations required for crystallisation, it is clear that peptide-peptide interactions forming the fibre are more favourable than peptide-hydration interactions, resulting in this specific aggregation/polymerisation rather than forming a crystal. Aggregation of calcitonin analogues has been previously observed in CD experiments, with peptide concentrations in the μM range. This has led to the hypothesis that calcitonin analogues can form trimers (Moe & Kaiser, 1985) and tetramers (Green *et al.*, 1987). However the formation of these fibrous polymers has never before been reported or proposed. In the crystallisation trials the high concentrations of sCT have also resulted in the formation of surface skins and gels. The formation of these unusual precipitates can adequately be explained by the formation of amyloid fibrils with subsequent further aggregations of these fibrils into the skins and gels. While the peptide micelles are clearly composed of fibrous material as seen in plate 4.8, the

fibrous nature of the other types of precipitate have not been determined. It has been shown that amorphous precipitate is not fibrous as shown in plate 4.10. Therefore, alternative possibilities for the origin of these other precipitates cannot be ruled out.

It is possible that the surface skins (plate 4.1b) occur as a result of the surface active properties of amphipathic peptides. Model peptide studies have shown that an amphipathic α-helix can be formed by an air-water interface (Kaiser & Kezdy, 1983). It has also been shown that amphipathic α-helices can form stable monolayers at the air-water interface (de Grado *et al.*, 1981). Thus these skins could be formed by the aggregation of these α-helices at the air-water interface. The formation of such skins could affect the water equilibration in the hanging drop experiments if they occurred before supersaturation. The fibrils of the peptide micelle are by their size, clearly composed of many of the single fibrils seen with the negative staining technique plate 4.4 & 4.5. From the S.E.M. micrographs the diameter of these larger fibrils must represent at least 25 of the individual fibres (as shown in plate 4.4) across the diameter. The nature of the interaction of the individual fibres is unknown, but the aggregation of amyloid fibrils into 'ribbons' has been previously observed (Fraser *et al.*, 1991). These are very similar to those observed in plate 4.5 but are not composed of anywhere near 25 individual fibres.

Despite providing some useful information on the solubility of sCT, the slow cooling method has not proved a good method for the screening of a wide range of crystallisation conditions. It has been difficult to set up and the rate of cooling has been difficult to control. Interpretation of the results has also not been

straightforward, but duplication of conditions in vapour diffusion experiments have helped assess the effect of individual parameters on the solubility of the peptide. There appears to be no specific effect of temperature other than increasing the rate of formation of these aggregates, while having no effect on their morphology. The slow cooling experiments have indicated a pH optimum for the formation of the fibrils around pH 6-7. Also at high concentrations (8-15 mM) aggregation occurs in the absence of precipitating agents. There also appears to be no specific role for divalent metal ions such as Ca^{2+} and Zn^{2+} in the aggregation process indicating no role for divalent metal ions in the formation of the fibrils themselves. The polypeptide has been shown to be particularily soluble in a variety of organic solvents (Table 4.3) which even at high peptide concentrations does not result in the formation of any aggregates. Solubility in this two component system could indicate that hydrophobic interactions are of importance in the solubility of this peptide, excluding any pH effects. This could be of particular relevance to the peptide's ability to bind and solubilise phospholipid micelles. In the three component system (the buffer is neglected due to it's low concentration) where the third component was water, peptide micelles were frequently formed. Quite why the presence of water should induce aggregation is unclear. It is known that when exposed hydrophobic sidechains are placed in aqueous solution, a structure will collapse to a state which shields the hydrophobic groups from solvent. As in protein folding, this leads to the formation of an hydrophobic core. With such a small peptide it could not be envisaged that the molecule itself forms such a core. Rather it is likely that hydrophobic sidechain interactions could occur as a consequence of aggregation.

Model compound studies have suggested that pairwise hydrophobic interactions can even occur when non polar sidechains are separated by a layer of water (Wood & Thompson, 1990). Thus the hydrophobic interactions of the sidechains, even if partly solvated, could promote fibrillogenesis.

Another effect of hydration, identified in an NMR study (Meyer *et al.*, 1991), is the induction of cis-trans isomerisation in the peptide. It was observed that the addition of water to sCT in TFE induced cis-trans isomerism at Pro-23. While the structure in TFE is an amphipathic α-helix, cis-trans isomerism at this residue has been observed in an amphipathic ß-sheet structure (Motta *et al.*, 1989). In this NMR study, carried out in cryoprotective DMSO-water mixtures, an amphipathic sheet structure from Leu-12 to Lys-18 was also observed. It was proposed that this structure was responsible for the initial interaction with lipids in the same way that signal sequences are thought to interact with bilayers (Briggs *et al.*, 1986). The existence of cis peptides has been identified in bombolittin (Bairakarti *et al.*, 1990), a bee venom peptide which can form an amphipathic helix in the presence of SDS. Cis-peptide conformation has also been identified by solid state NMR in the amyloid deposits of Alzheimer's disease (Spencer *et al.*, 1991) in which the peptide was also shown to exist in a single conformation. In both cases the stabilisation of the energetically unfavorable conformation (free energy difference of ca. 3kcal/mol at 298 K) has been proposed to occur as a consequence of the resulting newly formed interactions providing the necessary energy. In the case of bombolittin this is a peptide-micelle interaction, in the amyloid deposits of AD it is believed the cis bond is stabilised by the intermolecular interaction energy as it aggregates in the

formation of fibrils. However, the lack of formation of either peptide micelles or gels in the slow cooling experiment with sCT in a 50/50 mixture of TFE and water does not support the possible role of cis-trans isomerism in the formation of these amyloid fibrils. The hydrophobic and the cis-trans isomerisation effects are two of many possible consequences of hydration. While it is possible that both effects occur, it cannot be ruled out that there are other as yet undetermined hydration effects which have a role in the formation of these amyloid fibrils.

The hanging drop experiments have indicated an increased solubility at acidic pH's at calcitonin concentrations up to 7.5mM, with a solubility minimum around 6.5-7.2. The estimated isoelectric point for sCT is 10.4 (Maier *et al.*, 1977). Thus at a pH of 7 the peptide should carry an overall charge of approximately +3. These overtly positively charged molecules would tend to repel each other rather than aggregate. The formation of fibrils by synthetic peptides of A4/ß protein of AD have been shown to be similarily pH dependent (Fraser *et al.*, 1991). It was suggested that this process might be an isoelectric phenomenon since at the pH's at which fibrils were obtained (pH 3-8) the peptide had no charge. If sCT forms fibrils in a similar way to A4/ß, the process cannot be isoelectric.

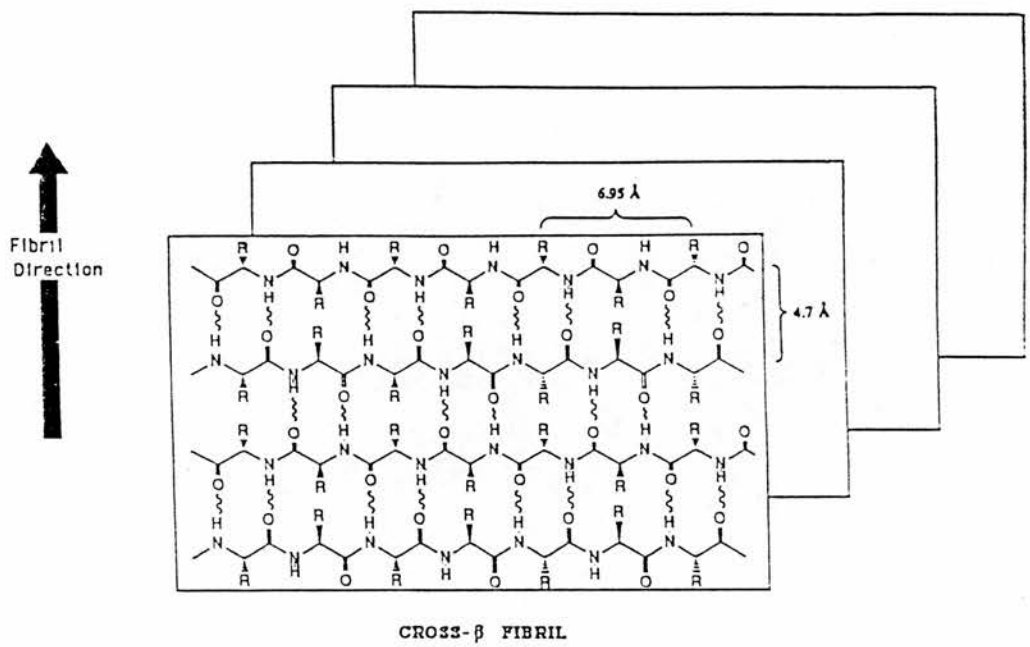
The alternative explanation proposed for the pH dependence of A4/ß fibrillogenesis was that it was dependent on one or more His-Asp/Glu salt bridges (Spencer *et al.*, 1991). Although the pH dependence of fibrillogenesis for sCT is different, it would indicate that certain charged residues have a role in fibrillogenesis. It is also possible that the ring structure of the amino terminus plays some role. No micellar aggregates were obtained with either of the two derivatives.

It has been proposed from NMR work (Meadows *et al.*, 1991) that this ring structure may act as a template upon which secondary structures may develop. That the vapour diffusion experiments with the polyethylene glycols and MPD did not lead to the formation of these amyloid aggregates, again indicates the importance of hydrophobic interactions. Peptide-solvent interactions in this case are presumably of similar magnitude to peptide-peptide interactions associated with fibrillogenesis. However in the presence of the helix inducing phospholipid DMPG, amyloid precipitate was obtained in PEG 4000 which had previously produced only amorphous precipitate. Since the use of additives in crystallisation trials is to improve crystal quality, the lack of success with the four additives could be expected. Octyl-glucoside and DMPG both affect the form of the precipitate with peptide micelles being obtained in PEG 4000 with only DMPG. It is highly likely that the surface skins obtained with octyl-glucoside are also fibrous since similar skins have been observed with peptide fragments of A4/ß amyloid protein of Alzheimer's disease (Burdick *et al.*, 1992) and shown to stain with Congo Red. Unfortunately it has not been possible to stain histochemically the skins obtained with sCT due to their physical fragility. In contrast to DMPG and octyl-glucoside, SDS which like DMPG induces helicity in sCT, actually increases the solubilty of sCT and leads only to amorphous precipitate. Quite why such opposing results have been obtained with these additives cannot be explained.

Since fibrillogenesis has been observed only in aqueous solutions, where sCT has been shown to possess no secondary structural features (Wuthrich, 1976), the necessary structure required for fibrillogenesis must be induced at the fibril itself. This would not therefore be

expected to be a rapid process. Rather it would be similar to the formation of a crystal where the slow approach to supersaturation allows molecules sufficient time to orient themselves for specific contacts to be made. This could explain why at high precipitant levels only non-specifc aggregates occur. Here, the rapid loss of solvation results only in amorphous precipitation. The induction of a structure at the fibre is further supported by the fact that fibrillogenesis can occur in the presence of helix promoting solvents such as DMPG. There has been no evidence that amyloid fibrils possess helical structure, rather all the structural studies indicate some form of ß-structure as proposed by Glenner (1980). Some structural rearrangement must therefore occurr during the fibrillogenesis. Unfortunately the techniques which have been used to characterise amyloid have provided no detailed structural information on the plaque itself. Only speculative models similar to Pauling's cross ß-structure have been proposed as shown in figure 4.8. Since amyloid forming protein sequences are so diverse it could not be imagined that one single model of the fibril could be applied to all of them. Thus it has been proposed (Lansbury, 1992) that the defining properties of amyloid characterise a class of related structures, of which the cross-ß fibril (Marsh *et al.*, 1955) is one member rather than the definitive structure. It is possible that the ß-sheet structure identified by NMR in DMSO-water mixtures (Motta *et al.*, 1989) could resemble the conformation in the fibre.

Figure 4.8 Schematic illustration of Pauling's cross-β fibril model (Marsh et al., 1955). This model is generally accepted to represent all amyloid. The anti-parallel arrangement of the peptide chains and the resultant hydrogen bonding network are shown for the front sheet.



Chapter 5 Crystallographic refinement of equine phospholipase A2

Chapter 5

5.1 Introduction

This chapter explains how two methods of crystallographic refinement were successfully applied to equine phospholipase A2 and the putative co-crystal of equine PLA-2 with an inhibitor compound crystallised as described in Chapter 4.

5.2 Crystallographic refinement using XPLOR

As previously explained in chapter 2, the purpose of any refinement process is to produce a model which agrees as closely as possible to the observed data but still has sensible geometry. Thus any refinement must produce the best fit between the observed structure factors F_O and those calculated from some model F_C. Thus F_C is a function of the x, y, z coordinate positions and B (temperature factor) for each atom of the model. However unlike small molecule crystallography, constraints or restraints are needed for proteins in order that the number of observations remains greater than the number of parameters of the model which is being refined. A variety of methods have been developed for protein crystallography which can do this. The relative merits of these methods make some more applicable in certain situations. The majority of the methods are least squares in essence, one of which will be discussed later. A relatively recent addition to the methods available is the use of molecular dynamic simulations. These are used to search conformational space, while at the same time trying to minimise the difference between the observed and calculated structure factors. This technique has proved to be a particularly powerful technique in the crystallographic refinement of protein structures (Brunger *et al.*, 1987, Brunger, 1988a). This has

subsequently been developed by Brunger into a refinement package called XPLOR (Brunger, 1988b).

The package incorporates simulated annealing (SA) (Kirkpatrick *et al.*, 1983), which allows exploration of the phase space of a complicated multi-parameter target function. In the crystallographic refinement this target function can be described

$$E_{TOTAL} = E_{EMPIRICAL} + E_{X\text{-}RAY}$$

where $E_{EMPIRICAL}$ is the empirical energy of the model and $E_{X\text{-}RAY}$ is the squared difference between F_O and F_C summed over all the reflections. The reason this method proves so powerful over conventional least squares methods is that energetically unfavorable 'uphill steps' can be taken across barriers between energy minima, with a probability $e^{-Etotal/k_BT}$, using an effective temperature as a control parameter (k_B is the Boltzmann constant). Conventional least squares can only take energetically 'downhill' steps and as a result can end up trapped in local minima. This is due to the limited radius of convergence, as restrained least squares refinement does not usually correct atoms whose positions are more than 1 Å in error. This necessitates frequent manual inspection and refitting using interactive computer graphics which is a time consuming process. The larger radius of convergence of XPLOR reduces the need for manual corrections. The name of the method is derived from the physical analogy of changing an amorphous glass (high energy, only local order) to a crystalline solid (low energy, long range order): the glass must first be liquefied, then slowly cooled or 'annealed' into the crystalline state. Slow cooling is required to prevent the system from becoming trapped in a metastable solid state with only local order. The restrained molecular dynamic simulation does not correspond to any

physical reality; it is simply a tool for carrying out a search of configuration space.

5.2.1 The XPLOR energy function

As mentioned earlier there are two components of the energy function (when performing crystallographic refinement) and these are:

$$E_{TOTAL} = E_{EMPIRICAL} + E_{X\text{-}RAY}$$

The first term $E_{EMPIRICAL}$ describing the energy of the molecule can be broken down further. Since it describes the interactions that stabilise the protein structure, the term can be replaced by a series of expressions that give the potential energy of the molecule as a function of its atomic positions. In XPLOR, the empirical energy function has the from:

$$E_{EMPIRICAL} = E_{BONDS} + E_{ANGLES} + E_{DIHEDRAL} + E_{IMPROPS} + E_{NONBONDED} + E_{SBOUNDS}$$

Each bond is treated as a simple spring obeying Hooke's law with a characteristic force constant and equilibrium bond length:

$$E_{BONDS} = k_b(r - r_o)^2$$

the second term accounts for the deformation energy of angles between the covalent bonds to a given atom:

$$E_{ANGLE} = k_\Theta(\Theta - \Theta_o)^2$$

The next two terms represent the intrinsic deformation energy for twisting about an axis through covalently bonded atoms:

$$E_{DIHEDRAL} = E_{IMPROPER} = k(1 + \cos(n\phi+\delta)) \text{ if } n>0$$
$$k(\phi - \delta^2) \quad \text{if } n=0$$

[n=periodicity δ=phase shift angle]

The reader is directed to Levitt (1982) for a fuller explanation of these two angular terms and an explanation of the difference in their definition. Because a harmonic approximation (Jack & Levitt, 1978) is used for the deformations in bond lengths and angles, the number of times the direct computation of structure factors and their derivatives have to be calculated is reduced.

The nonbonded energy term has the form:

$$E_{NONBONDED} = E_{VDW} + E_{ELEC} + E_{PVDW} + E_{PELEC}$$

These correspond to the van der Waals and Coulombic interactions of the atoms and those related through symmetry operations. No explicit hydrogen bonding term is included since the hydrogen bonding energy can be accounted for satisfactorily by electrostatic and van der Waals terms. The parameters for all these terms have been taken from CHARMM (Brooks *et al.*, 1983) and have been previously modified to avoid artefacts of the high temperatures encountered during simulated annealing.

5.2.2 Crystallographic refinement

The second term $E_{X\text{-}RAY}$ is described thus:

$$E_{X\text{-}RAY} = E_{AX\text{-}RAY} + E_{PX\text{-}RAY}$$

The second term, $E_{PX\text{-}RAY}$, describes phase information if present. No phase information was used in the refinement of equine PLA-2 since the initial model was obtained from the direct method of molecular replacement. The first term, $E_{AX\text{-}RAY}$ consists of the weighted differences between observed and calculated structure factor amplitudes,

$$E_{AX\text{-}RAY} = W_A/N_A \, \Sigma(hkl) \, W_{(hkl)}[\mid F_{obs}\,(hkl) \mid - k \mid F_c\,(hkl) \mid]^2$$

where (hkl) are indices referring to the reciprocal lattice points of the crystal and extends over all observed and selected reflections with indices (hkl).The overall weight W_A relates the crystallographic term to the empirical energy function. The normalisation factor N_A allows the weight W_A to be approximately independent of the resolution range used during SA-refinement. N_A has been set to:

$$N_A = \Sigma(hkl) \, W_{(hkl)} \mid F_{obs}\,(hkl) \mid^2$$

The necessary constants used to approximate the atomic scattering factor are from the International Tables for Crystallography Volume 4 (1974). The structure factors are computed using the Fast Fourier Transformation (FFT) method (Agarwal, 1978). In order to minimise the computational requirements of SA-refinement an approximation is included. At every dynamics or minimisation step the $F_{calc}(hkl)$ and it's first derivative are only recalculated if the coordinates of the model have changed by a specified amount, for example 0.2 Å during molecular dynamics and 0 - 0.05 for minimisation.

5.2.3 Molecular dynamics

Having a knowledge of the energy of the system as a function of atomic coordinates, allows one to perform molecular dynamic simulations. There are forces acting on the atoms of the system. These forces, are derived from the empirical potential energy that describes the stereochemical and nonbonded interactions. These can be used to calculate the dynamic behaviour of the system by solving Newton's equations of motion for the atoms as a function of time. The atoms are assigned initial velocities with a Maxwellian distribution at some given temperature. The temperature can be measured by the mean kinetic velocity of the atoms

$$\tfrac{1}{2} \Sigma m_i \langle v_i^2 \rangle = 3/2 \, N k_B T$$

where N is the number of atoms in the system, $\langle v_i^2 \rangle$ is the average velocity squared of the ith atom and k_B is the Boltzmann constant. A simulation is performed by finding the acceleration a_i of atom i from Newton's law $F_i = m_i a_i$. From the acceleration of the atom it is possible to calculate it's new position r_i after a time interval $t+\delta t$, given r_i at time t:

$$r_i(t+\delta) = r_i(t) + v_i \delta t + \tfrac{1}{2} a_i (\delta t)^2 + ..$$

The quantity δt must be very small so that the potential energy does not change too much during each time step; in XPLOR this is about one femtosecond (10^{-15}s). The desired temperature is maintained by coupling the dynamics to an external heat bath via frictional dampening as described by Berendsen *et al.*, (1984). This method is

a mathematical abstraction and has no physical reality, but the algorithm is particularly suited to the simulated annealing technique.

5.2.4 XPLOR refinement of equine PLA-2.

Simulated annealing has been applied to the refinement of equine PLA-2. The relevant crystallographic details are described in table 5.1. The initial data set was collected by Dr. M. Walkinshaw on a CAD-4 in Delft, Holland and a second data set was collected on a FAST area detector in Basle. The CAD-4 data set was used in the initial 'solving' of the structure. Molecular replacement was used by Dr. Walkinshaw using the program MERLOT (Fitzgerald, 1988) to locate the structure within the unit cell.

During a crystallographic experiment only the relative intensities of reflections can be measured and not their phases. This phase information is lost during the experiment. However, once the approximate positions and identities of all the atoms in the unit cell are known, the amplitudes and the phases of the structure factors can be readily calculated. The method of molecular replacement involves finding these approximate positions by the use of an already determined structure which is believed to be sufficiently similar to the new structure. It is generally the case that proteins whose sequences are similar tend to have a similar overall fold. In this case, the crystal structure of Bovine PLA-2 of Dijkstra *et al.*, (1981) was used. The bovine sequence shares a high sequence homology with that of the equine (76.8% identity). This bovine structure was determined at 1.7 Å resolution and has been refined to a crystallographic residual of 0.171 (Dijkstra *et al.*, 1981) making it the ideal candidate for a search molecule. The necessary changes to

Table 5.1 Crystallographic details for the refinement of equine PLA-2.

Crystallised in:	30% Ammonium sulphate/6mM $CaCl_2$/100mM Imidazole-HCl, pH 7.2.
Structure solution:	Molecular replacement
Space group:	C2 2 molecules/asymmetric unit each of 125 residues.
Unit cell parameters:	a=83.2, b=45.44, c=65.13 Å $\alpha=\gamma=90.0^{\circ}$, $\beta=101.98^{\circ}$

Table 5.2 Analysis of the CAD-4 data as a function of resolution between 8.0 and 2.45 Å.

Resolution Range (Å)	Number of reflections	R-factor
2.45 2.56	266	0.5127
2.56 2.69	476	0.5173
2.69 2.85	498	0.4061
2.85 3.06	723	0.3540
3.06 3.34	1072	0.3145
3.34 3.78	1096	0.2865
3.78 4.61	1086	0.2578
4.61 8.00	1107	0.2681

the bovine structure, in compliance with the equine sequence, were carried out by Dr. Walkinshaw using the molecular modelling program SYBYL (Tripos associates, 1991) prior to the molecular replacement.

5.2.5 Simulated annealing refinement of equine PLA-2

The initial data set used for the refinement extended to 2.45 Å resolution. The resolution range used in the initial refinement was from 8 - 2.7 Å. This data set contained 5537 unique reflections. An analysis of the data as a function of resolution is shown between 8.0 and 2.45 Å in table 5.2.

The first step in the refinement was the determination of the ideal weight W_A between the structure factor information and the empirical energy function. This was calculated by performing a brief molecular dynamics calculation without the structure factor information and then comparing the norm of the gradient of the empirical energy function and the norm of the gradient of the structure factor information. This weight was used in all subsequent parts of the refinement carried out with this data. The initial R-factor was 45.0%. Since the initial structure was essentially that of an already highly refined crystal structure only 40 cycles of energy minimisation were carried out. No harmonic constraints on the C_α positions were used and the R-Factor after this stage was 44.8%. Next an overall B-Factor refinement was carried out using an 8 step least squares optimisation. An overall B-Factor of 25.87 was calculated for all non-hydrogen atoms. This structure was then run through a simulated annealing refinement "slow cooling" protocol as shown in figure 5.1. The starting temperature was 3000K and the system was coupled to a heat bath. The temperature was cooled by 50K every 50 steps of the dynamic simulation with a 0.0005

Figure 5.1 Slow cooling protocol for equine PLA-2 using XPLOR.

Methodology

1. Prepare: 40 cycles of Energy Minimisation
2. B-overall refinement of all non-H atoms.
3. Slow cooling from 3000 K to 300 K in 50 K steps, Timestep 0.0005ps
   50 integration steps.

Simulated Annealing Schedule

```
remark 3000 K to 300 K in 50 K steps
resolution-limits= 8.0   2.7
tolerance= 0.2                    {Use linear approximation until}
                                  {any atom has moved by more    }
                                  {than 0.2 A                    }
wa=50000                          {These are the weights from    }
wp=0.0                            {weights.inp                   }

vector do (vx=maxwell (3000.0)) (all)
vector do (vy=maxwell (3000.0)) (all)
vector do (vz=maxwell (3000.0)) (all)
vector do (fbeta=100) (all)

evaluate ($1=3000.0)
while ($1> 300.0) loop main
dynamics verlet                   {Invoke verlet integration}
timestep=0.0005                   {Small timestep since the}
                                  {temperature is high      }

nstep=50                          {Number of integration steps}
iasvel=current
tcoupling=true tbath=$1           {Berendsen method, heatbath}
nprint=25 ipfrq=0                 {Output Control            }
 end
evaluate ($1=$1-50)
end loop main
```

Results: Initial R Factor= 45.0 After B-overall refinement=35.6
Final R Factor= 28.8

picosecond timestep. The final temperature was 300K. The R-Factor after this stage was 28.8. Figure 5.2 outlines the comparison of the C_α's of the initial (EPLA1) and final structures (EPLA_SA1) for both the molecules A and B which form the asymmetric unit. The rms difference between the starting (EPLA1) and final (EPLA_SA1) structure was 0.903 Å for the main chain atoms and 1.722 Å for the sidechain atoms. Figure 5.3 outlines the difference in C_α positions between both molecules of the asymmetric unit cell after optimal superposition of A and B using SYBYL (Tripos Associates, 1991). The rms difference between the main chains of molecule A and B was 0.804 Å. Table 5.3 shows the stereochemical analysis of EPLA_SA1 indicating the larger movements of both the backbone and the sidechains. Both a $2F_O$-F_C map and omit maps with coefficients $2F_O$-F_C were used to remodel atoms back into density. Less than 10% of the structure was omitted in any one omit map. Remodelling was carried out using FRODO (Jones, 1982). The main areas that required remodelling were around residues 63 - 67 and around the C-terminus in both molecules. There were also breaks in main chain density in a number of the loop regions. This structure (EPLA2) was then used as the starting model with the same protocol as before in a second "slow cooling" refinement again using the data between 8-2.7Å. The R factor after this annealing run was R=28.0% for EPLA_SA2. Once again remodelling using $2F_O$-F_C and omit maps were carried out to produce EPLA3. At this stage a higher resolution data set (here called FAST1) was collected by Dr. M. Walkinshaw on a FAST area detector in Sandoz, Basle. This data set extended to 2.0 Å and had a low R_{merge} (6%). However, after analysis of the higher resolution reflections a resolution cut off was applied and the resolution range used in the refinement was 15 - 2.5 Å which contained

Figure 5.2

Distance (in Angstroms) between the CA positions of EPLA1 and EPLA1_SA1. The solid line represents molecule A and the dotted line molecule B. These were distances were calculated after optimal superposition using SYBYL.

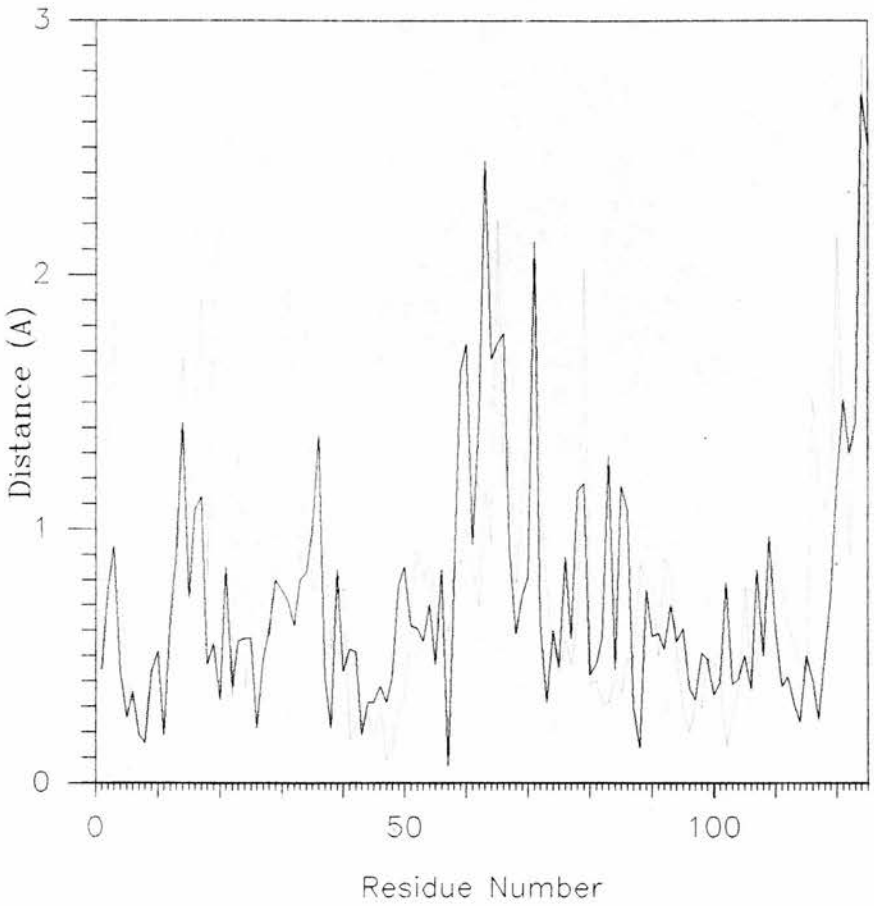


Figure 5.3
Distance (in Angstroms) between the CA positions of molecule A and molecule B of EPLA1_SA1. These distances were calculated after optimal superposition using SYBYL.

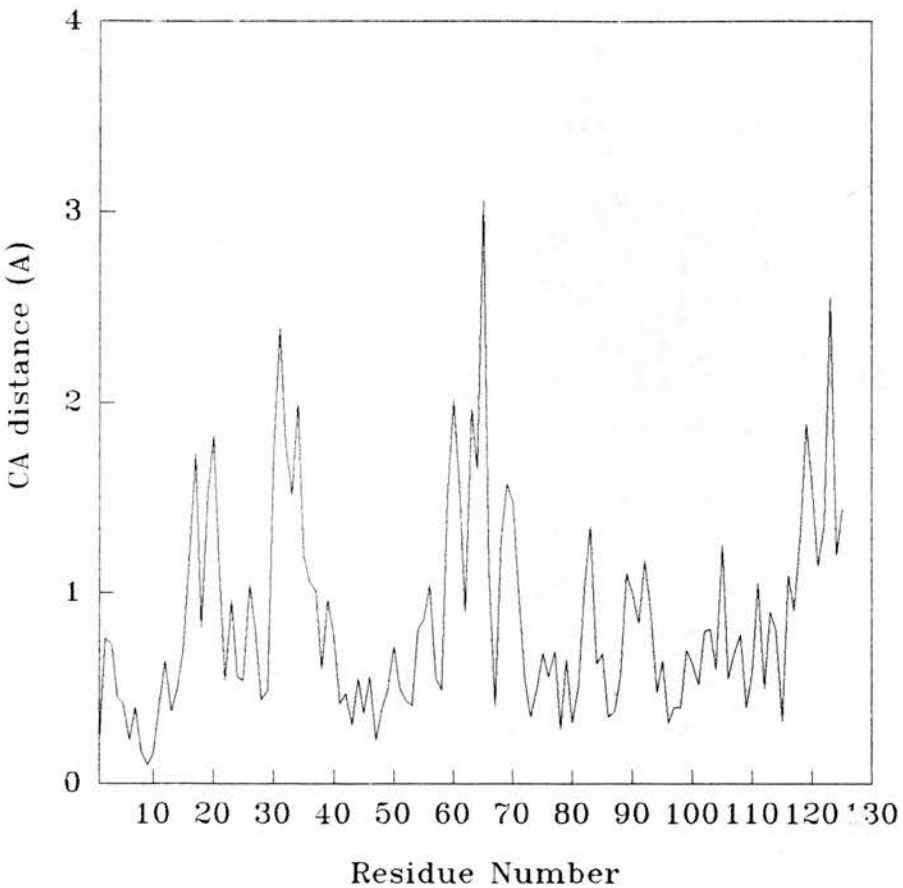


Table 5.3 Stereochemistry of EPLA1_SA1 illustrating deviations from ideality. Also shown are the largest sidechain and backbone movements (B elements) which occurred during the "slow cooling" refinement. These distances or B elements are given in Angstroms.

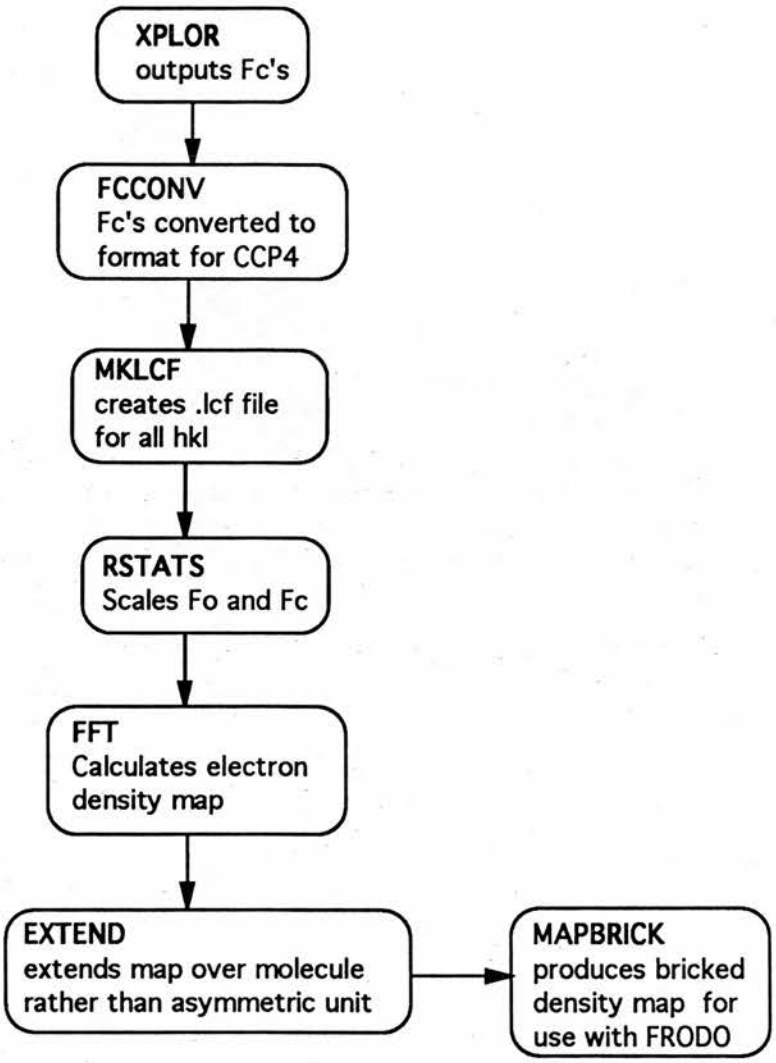
Bonds	12 bonds (sidechain) in A
deviation > 0.06	12 bonds (sidechain) in B
rms deviation	0.02
Angles	
deviation > 10	10 (main) 20 (sidechain) in A
	17 (main) 17 (sidechain) in B
rms deviation	4.01
Comparison to EPLA1	
Backbone	
rms	0.9
atoms B element > 2.0	10 atoms (5 residues) in A
	15 atoms (11 residues) in B
largest B element	Lys124 CA in A (2.8 A)
Sidechain (not Hydrogen)	
rms	1.72
atoms B element > 5.0	4 atoms (3 residues) in A
	7 atoms (5 residues) in B
largest B element	Tyr 19 OH in A (7.07 A)

7069 unique reflections. This data set was used with the EPLA3 structure. A new ideal weight was calculated for this data set and found to be 0.1132×10^6. This structure was then used with the same "slow cooling" protocol used previously to produce EPLA-SA3. The final R-factor at this stage was 29.8%. The electron density maps with this new data were significantly better and more featured. However the C-terminus was still badly defined and some of the loop regions still had breaks in the main chain density. The most notable of which was around residue 65 in both molecules. The omit maps did not indicate alternative chain tracing either in this region or in the C-terminus. Another round of remodelling produced EPLA4 which was used in an altered annealing protocol. As before the starting temperature was 3000K and the system was coupled to a heat bath. However, the temperature was now cooled by 25K every 50 steps of the dynamic simulation with a 0.0005 picosecond timestep. The final temperature was 300K. The final R-factor dropped only to 29.7% to produce EPLA_SA4. Since very little improvement in the structure was occurring with subsequent computationally expensive XPLOR runs, the refinement using XPLOR was halted at this stage.

5.2.6 Generation of electron density maps

The electron density maps were not calculated using XPLOR but from the CCP4 suite of programs as outlined in figure 5.4. These maps were normally $2F_O$-F_C maps but often "omit" maps were produced. Parts of the model were omitted (by setting the occupancy of the omitted atoms to zero) in the calculation of structure factors. The linear length of the structure omitted was never more than 10%. This has been suggested to provide a less model-biased electron density map and was thought to

Figure 5.4 Method of generating electron density maps during the initial refinement of equine PLA-2 with XPLOR.



help with the manual refitting of the structure into the electron density.

5.2.7 Discussion

XPLOR has proved very successful in the initial stages of refinement. The slow cooling methodology on it's own seems sufficient in producing a reasonable final model. No equilibration stage was used and the final structure had reasonable stereochemistry indicating that a final minimisation stage would not be necessary. It was noticed that some lysine sidechains, in particular Lys 124 adopted peculiar conformations and formed hydrogen bonds with the backbone carbonyl oxygens. However, the electron density of the $2F_O$-F_C and "omit" maps allowed these to be easily remodelled into density with sensible geometry. In subsequent refinements charged sidechains had their charges switched off. The successful progress of the initial refinement indicated both that the structure was successfully located within the unit cell and that the starting model was adequate. From table 5.3 it is clear that the largest movements in the starting structure EPLA were obtained in the side chain atoms. The largest of these movements occurred in residues which had been substituted from the sequence of bovine PLA-2 to comply with the equine sequence. A good example of this is Tyr 19B OH which moved over 7 Å during the refinement. In the bovine sequence this residue is a Leucine. A movement of this scale would only be possible, manually, with molecular graphics. However, not all of these large movements are by substituted side chains. A good example of this is provided by the CD1 atom of Leu 31B which has shifted 5.5 Å. These provided early indications that the equine structure and the bovine structure were

slightly different in overall conformation and quite different at certain areas within the sequence. Comparison of the C_α's of the starting and final structures in figure 5.2 indicates four major areas in the sequence where large rms differences occur. These areas are located particularly in the loop region between helix A (1-14) and helix B (17-22) and the calcium binding loop (25-37). The worst two areas occur between helix D (54-58) and helix E (89-108) and the area of the last 6 residues at the C-terminus. These areas are consistent in both molecules although it appears that molecule B is slightly further from the starting structure than molecule A.

5.3 Crystallographic refinement using TNT

It has become clear that while refinement methods incorporating molecular dynamics simulations are very powerful, they are most productively used at the early stages of structure refinement (Jhoti, 1989). Subsequent refinement steps are completed with the more conventional least-squares type refinement methods.

Further progress in the refinement process requires the use of a least squares refinement program. This method is less computationally intensive than methods which incorporate dynamical simulations. Least squares methods can only take steps that are energetically favourable as the function is minimised. Since the method has a small radius of convergence the structure can easily be refined into local minima. This is then overcome by manual inspection and remodelling using molecular graphics.

TNT is such a restrained least squares refinement package (Tronrud *et al.*, 1987) and has been applied to the refinement of equine

phospholipase A2. The function to be minimised in TNT is

$$M = \Sigma_{(hkl)} W_{(hkl)} [F_o(hkl) - F_c(hkl,p)]^2$$

where $F_o(hkl)$ is the experimental value for observation hkl, $F_c(hkl,p)$ is a corresponding value calculated from the coordinate and thermal parameters p that specify the structural model and W(hkl) is the desired weighting function. This is performed by summation over all the independent reflections. To obtain the best fit the parameters p are considered as variables which may be adjusted to minimise M. This is done by setting the derivatives of M with respect to each of these parameters p equal to zero. This leads to p independent simultaneous equations. These equations are not linear in the parameters, since they involve trigonometric and exponential functions and so cannot be treated directly by least squares methods since the method requires linear equations. However, if a reasonable starting set of parameters is known, then it is possible to derive a set of linear equations in which the variables are the shifts from the starting parameters, rather than the parameters themselves. This is done by expanding in a Taylor's series about the trial parameters, retaining only the first-derivative terms on the assumption that the shifts needed are sufficiently small that the terms involving second derivative and higher order derivatives are negligible. Thus the process starts with some initial set of parameters which are used to produce small shifts in these parameters and a new set is produced. This new set of parameters is used again in a new cycle of refinement. In TNT the function to be minimised is further separated into the sum of two terms based on the crystallographic observations s and the

stereochemical observations b.

$$M = \Sigma_s W(s)[F_o(s) - F_c(s,p)]^2 + \Sigma_b W(b)[F_o(b) - F_c(b,p)]^2$$

The gradient of M can also be separated into these terms. The stereochemical observations take the form of restraints and are specified as bond lengths, bond angles and so on. There are six classes of stereochemical information with which the structural model can be restrained: bond lengths, bond angles, torsion angles, trigonal planarity, general planarity and contact between non-bonded atoms. These restraints can be individually weighted during the process of refinement or modified to suit the desired requirements.

The crystallographic term which is minimised is

$$M = \Sigma_s W(s)[k \mid F_o(s) \mid - \mid F_c(s,p) \mid]^2$$

At the beginning of each cycle of refinement the scale factor k is determined by minimising

$$M(k,B) = \Sigma_s W(s)[k \mid F_o(s) \mid -\exp(-Bs^2/4) \mid F_c(s,p) \mid]^2$$

where $s = \sin\Theta/\lambda$ and $F_c(s)$ is treated as a constant. The thermal factor B allows for the initial overall discrepancy between the F_o and F_c's. As the refinement proceeds this B is absorbed within the thermal factors of the atoms and during successive cycles rapidly approaches zero. The structure factors are calculated by a space-group-specific fast Fourier transform (FFT) (Ten Eyck, 1977). The gradients are calculated using a modified version of the method of Agarwal (1978).

The modifications are described more comprehensively in Tronrud et al (1987). Essentially the gradients are calculated by first producing an (F_O-F_C) map for the molecular volume. For each parameter in the model a convolution, evaluated at the atomic position, is calculated between this F_O-F_C map and the derivative of the calculated atomic electron density function for the atom involved. Having determined the gradient for each parameter a shift is applied in the opposite direction. The size of the shift then has to be determined by a line search along the gradient vector. An overview of the TNT refinement process is shown in figure 5.5.

5.3.1 TNT refinement of equine PLA-2

The last model from the XPLOR runs EPLA_SA4 was remodelled on the graphics using FRODO to produce the starting model in the TNT refinement PLATNT1. At this stage another FAST data set was available (here called FAST2) and attempts were made to scale the two FAST data sets (FAST1 and FAST2) using the CCP4 program ANISOSC. Regardless of which data set was treated as the derivative, the two sets of data scaled together poorly (R_{scale}=18% using F^2). Analysis of the scaling as a function of $\sin\Theta/\lambda^2$ could not provide an obvious explanation for this. It was noted that of the three data sets available, the two FAST data sets scaled together the best, with the FAST1 and the CAD4 data, not surprisingly, scaling together the worst (39%). Rather than used badly scaled data the refinement was continued with this new FAST2 data set since it also had a very low R_{symm}[I] value of 4.1% and also

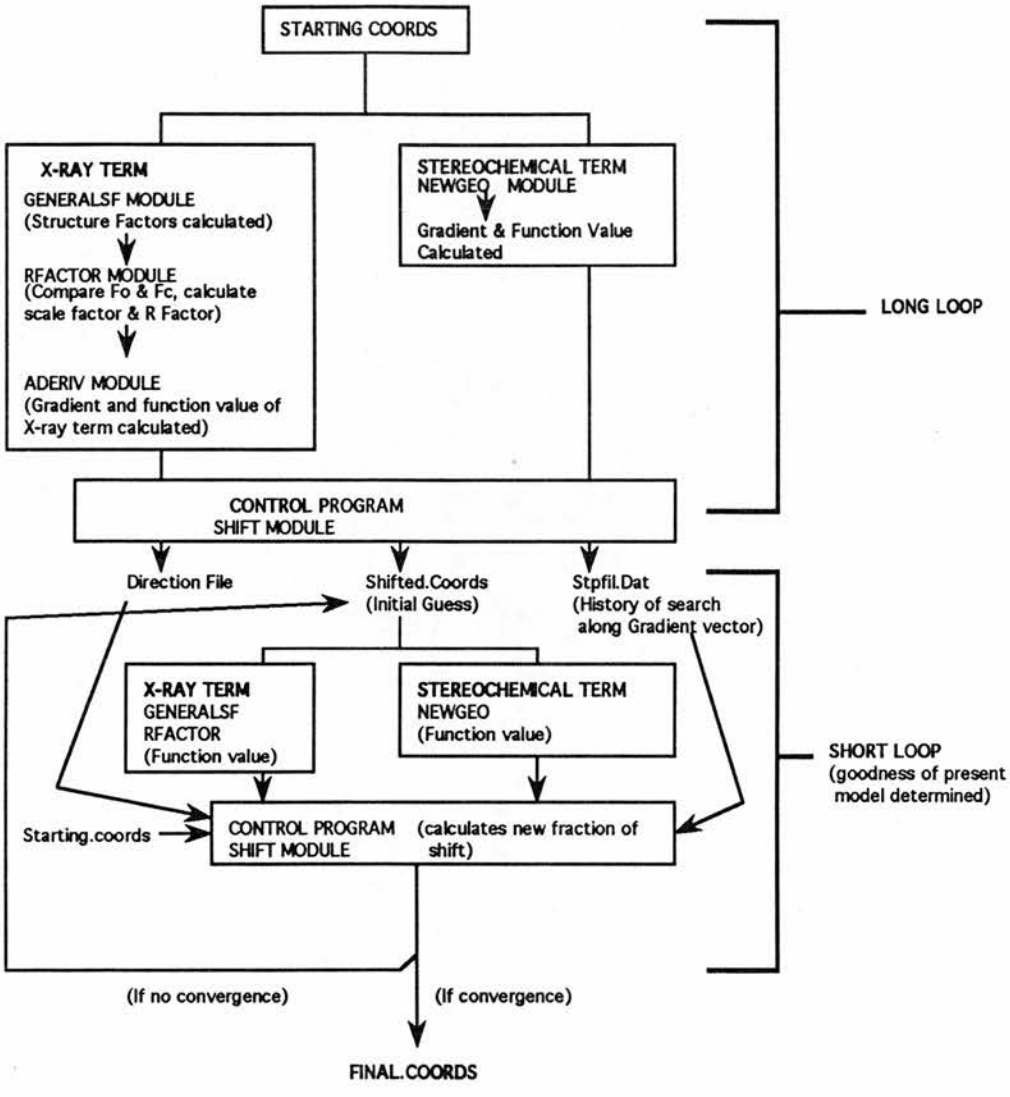
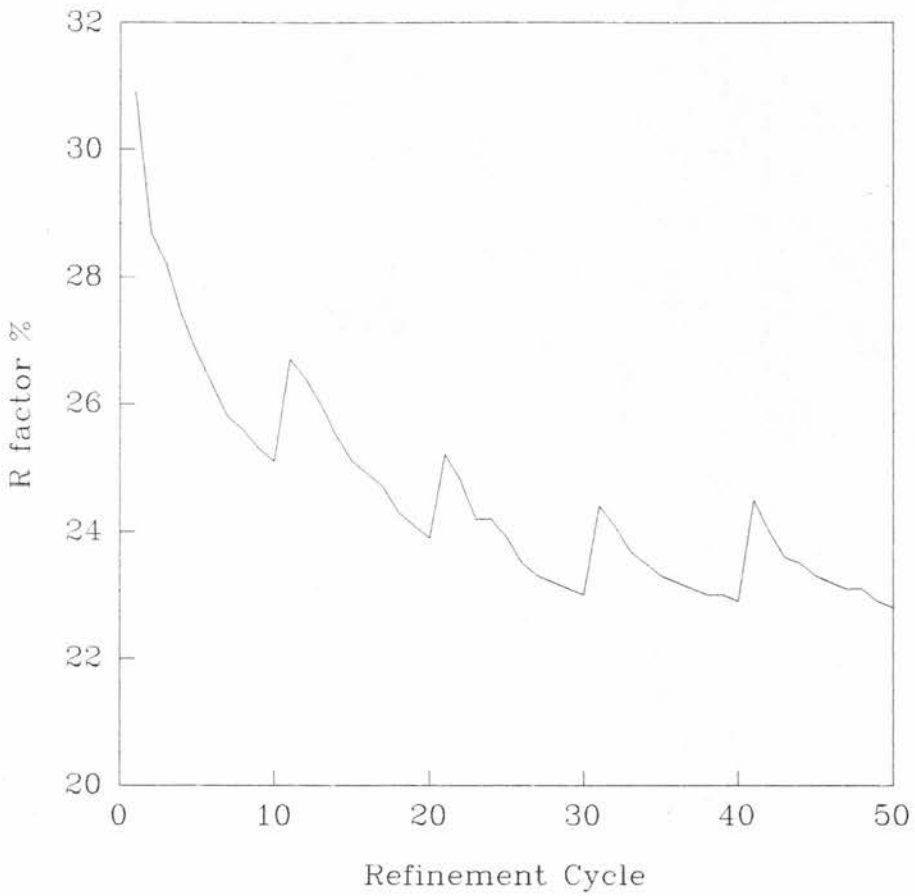


Figure 5.5 Flow diagram of a TNT refinement cycle.

extended to 2.0 Å. The data set used contained 13454 reflections of which 7969 were unique reflections and the resolution range for the refinement was 15-2.4 Å. This data was 82.6% complete to 2.4 Å. The diffraction data had to be first sorted with the index l varying the fastest and also h,k,l had to be transposed to -h,k,-l (in this space group these reflections are equivalent) for use in TNT. The refinement of phospholipase A2 was carried out in two stages. The first stage consisted of weighting up the crystallographic term. This resulted in the R-factor improving (getting smaller) at the expense of the geometry. Obtaining a suitable weighting scheme for this stage involved a mainly trial-and-error process. The refinement method used was the conjugate gradient (CG) method (Fletcher & Reeves, 1964). This method has a better rate of convergence than the steepest descent (SD) method. For good results with the CG method it was necessary firstly to run one cycle of SD and then at least three cycles of CG. The weights could not be adjusted between these cycles. Refinement cycles were followed up with temperature factor refinement before the structure was remodelled on the graphics and the whole process repeated. Temperature factors were refined between the range 0-100 $Å^2$. Modelling of the structure was accomplished using the program FRODO (Jones, 1982) on an Evans and Sutherland PS300 graphics machine. The geometry was not allowed to deteriorate too badly, however because there were no chiral restraints in this program three residues had adopted the wrong hand, residues Asp 66A, Glu 21B and Asp 66B. These were subsequently corrected using FRODO before further refinement cycles. The course of this initial stage of refinement is indicated in figure 5.6. The initial R-factor was 30.9% and after fifty cycles of CG refinement and four rounds of remodelling on the graphics, each

Figure 5.6 The course of the first stage of refinement using TNT.



followed by two cycles of temperature factor refinement, this had dropped to 23.1%. The second stage involved decreasing the crystallographic weight to restore the geometry of the model while trying to maintain the R-factor value. A round of refinement was started by carrying out one cycle of SD refinement followed by ten cycles of CG refinement. The round of refinement was completed with two cycles of temperature factor refinement, before the model was inspected on the graphics. At this stage water molecules were added where there was clear evidence of density both in $2F_O$-F_C (contoured at 1σ) and F_O-F_C (contoured at 2σ) with acceptable hydrogen bonding geometry. Water molecules with thermal factors greater than 50 $Å^2$ were deleted. The progress of this second stage of refinement is outlined in table 5.4.

5.3.2 Discussion

With the refinement of the structure completed, the final model contained a total of 1936 protein atoms, and 112 water positions had been included that made chemically reasonable contacts. The Ca^{2+} ions were refined with partial occupancies in the latter stages of the refinement. This was done because the quality of the electron density in this region, when contoured at 2σ, did not suggest a fully occupied site. Refinement of equine phospholipase A2 structure resulted in the reduction of the R-factor from 30.9% to 21.7% for data in the resolution range 15-2.4 Å, excluding reflections for which $I_{hkl} < 1\ \sigma(I)$. The rms shift of the atomic parameters in the last cycle of refinement was 0.008Å. The observed deviations in the geometric parameters of the final model are outlined in table 5.5. The stereochemistry of the final model is shown in figure 5.7 a-b and 5.8

Table 5.4 Summary of the latter stages of refinement of equine PLA-2.

Cycle	R-value	Comments
51-55	23.1 → 22.9	Structure after 50 cycles contained 3 residues out of hand A:66, B:21 and B:66. These were corrected on graphics before cycle 51.
56-65	24.1 → 22.6	Geometry weights increased (x10) for planarity terms and (x2) for the rest.
66-71	22.6 → 22.4	18 water positions added,11 sidechains repositioned.
72-84	24.3 → 22.6	One water removed, 6 moved back into density. 27 new water positions added, 3 sidechains moved.
85-98	24.2 → 21.3	Two waters removed 11 moved back onto density.
99-112	21.3 → 20.7	35 new water positions added. 3 waters moved back on to density.
113-126	21.4 → 20.9	40 new water positions added,5 sidechains moved.
127-140	22.4 → 20.6	9 waters removed 10 new water positions added.
141-168	21.2 → 20.9	7 waters moved back onto density
169-187	20.9 → 21.4	8 waters removed, 13 new water positions added. Calcium ions refined with partial occupancies. Geometric terms upweighted.
188-201	21.4 → 21.7	7 waters moved back on to density.

Table 5.5 Weighted deviations from 'ideal' geometry of the final model in the refinement of equine PLA-2.

Parameter	rms deviation	sigma
Bond length (A)	0.014	0.02
Bond angle (o)	2.894	3.0
Torsion angle (o)	14.879	15.0
Trigonal atom (A) non-planarity	0.019	0.02
Planar groups (A)	0.014	0.02
Bad contacts (A)	0.20	0.10

Figure 5.7 a. Ramachandran plot for A chain of final model from TNT. The glycine conformations are highlighted and Asp 66 which is also in an unfavourable conformation.

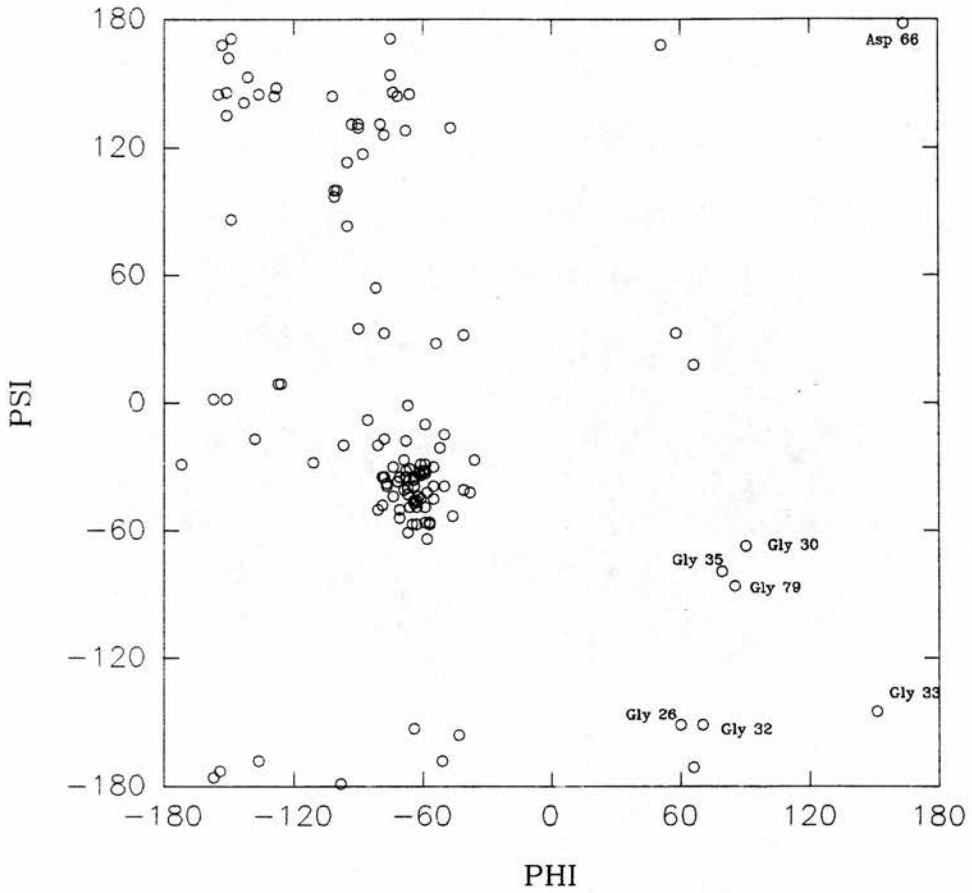


Figure 5.7 b. Ramachandran plot for B chain of final model from TNT. The glycine conformations are highlighted and Ser 34 Asp 66, Gln 67, Arg 122 which are in unfavourable conformations.

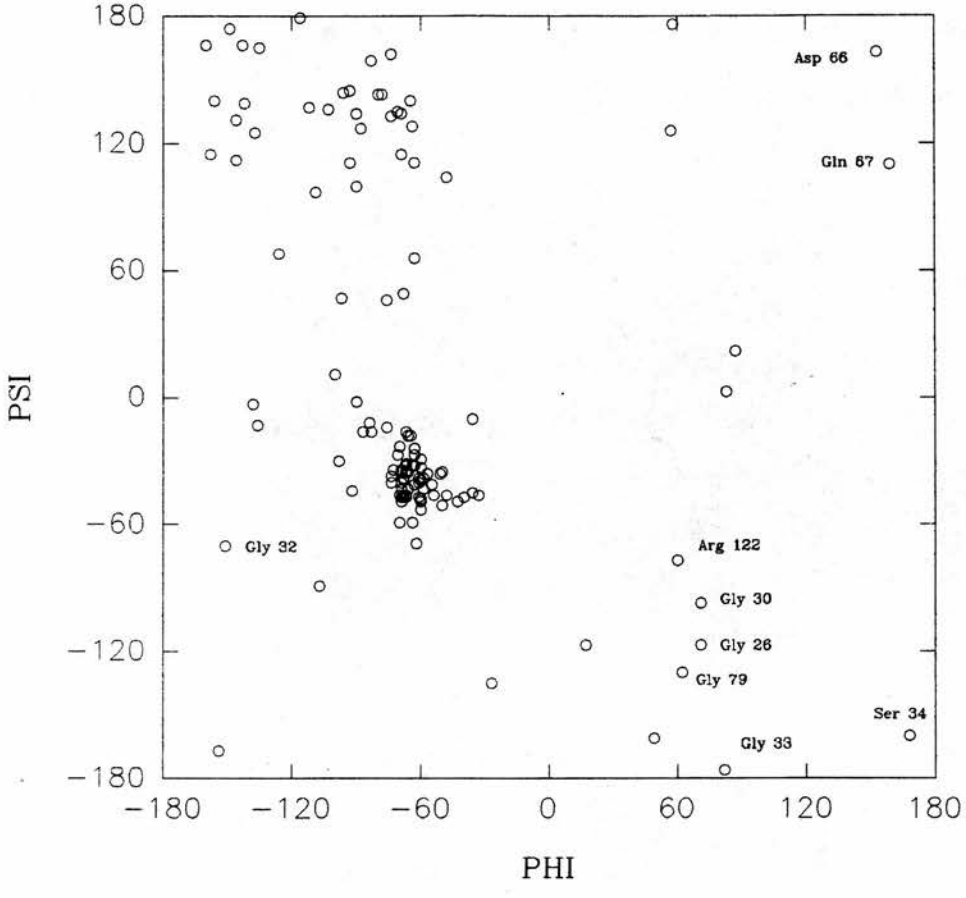


Figure 5.8 a. Omega plot for the A chain of the final model obtained from TNT.

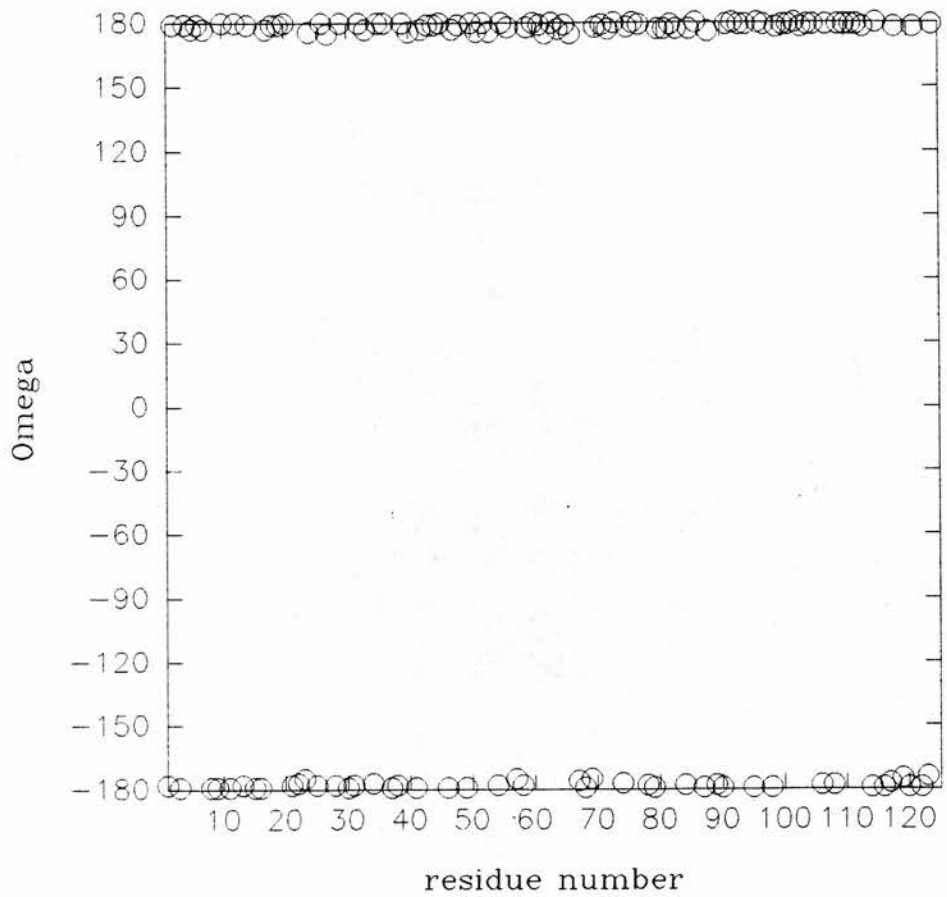
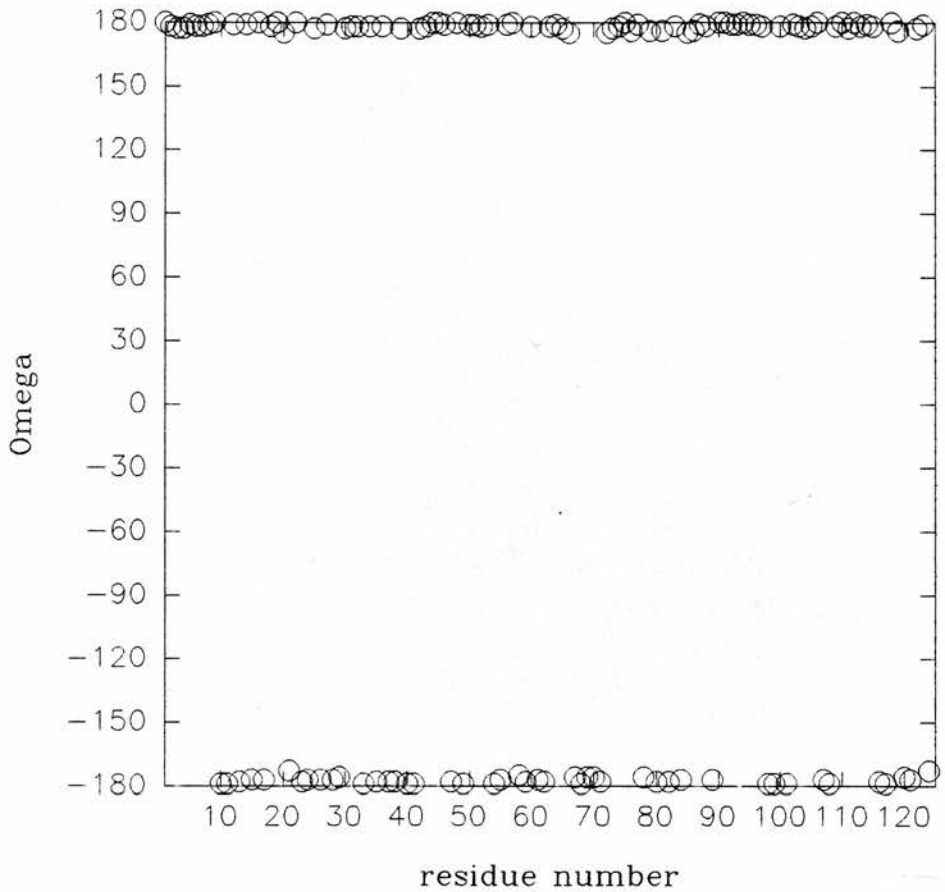


Figure 5.8 b. Omega plot for the B chain of the final model obtained from TNT.



a-b. The average B-factors for the main chains of the final model are shown in figure 5.9 and those of the sidechains in figure 5.10. The average B-factor of the solvent molecules is 48.3 $Å^2$.

While it is clear the structure has adequate geometry as indicated in table 5.5, figures 5.7a and 5.7b show the unusual geometry adopted by the glycine residues required for the Ca^{2+} binding loops of both molecules. Also figure 5.7b shows unfavorable conformations adopted by the region 66-67 in molecule B. Analysis of the thermal factors of these residues, figure 5.9 and 5.10 indicates that this region also has high thermal factors for both the main chain atoms and the sidechain atoms. This region of molecule B has poorly defined electron density which can explain these high values. Other areas of the structure where electron density is poor are also indicated by high thermal factors. There are three regions in particular that are obvious from figures 5.9 and 5.10. The regions, common to both molecules are between residues 28-35, 64-68 and 118-125. These are all loop regions in the structure. In molecule A there is an additional area with high thermal factors between 15-16 which is also a loop region immediately after the first helix. Both Ca^{2+} ions were refined with partial occupancies with values of 0.77 for the ion associated with molecule A and 0.92 for the ion associated with molecule B. The partial occupancies of these ions may explain why the binding loop is so badly defined. This flexible loop region being disordered in the absence of the Ca^{2+} ions. The binding constant for Ca^{2+} in the bovine enzyme is 4mM at pH 6 (Dutihl *et al.*, 1975) and it was assumed that the K_{Ca} for the equine enzyme would be similar. This work would suggest that in order to saturate all the sites in the enzyme a higher Ca^{2+} concentration be used in the crystallisation trials. Despite

Figure 5.9 average B-factor for main chain atoms of final model from TNT. Molecule A and molecule B are shown.

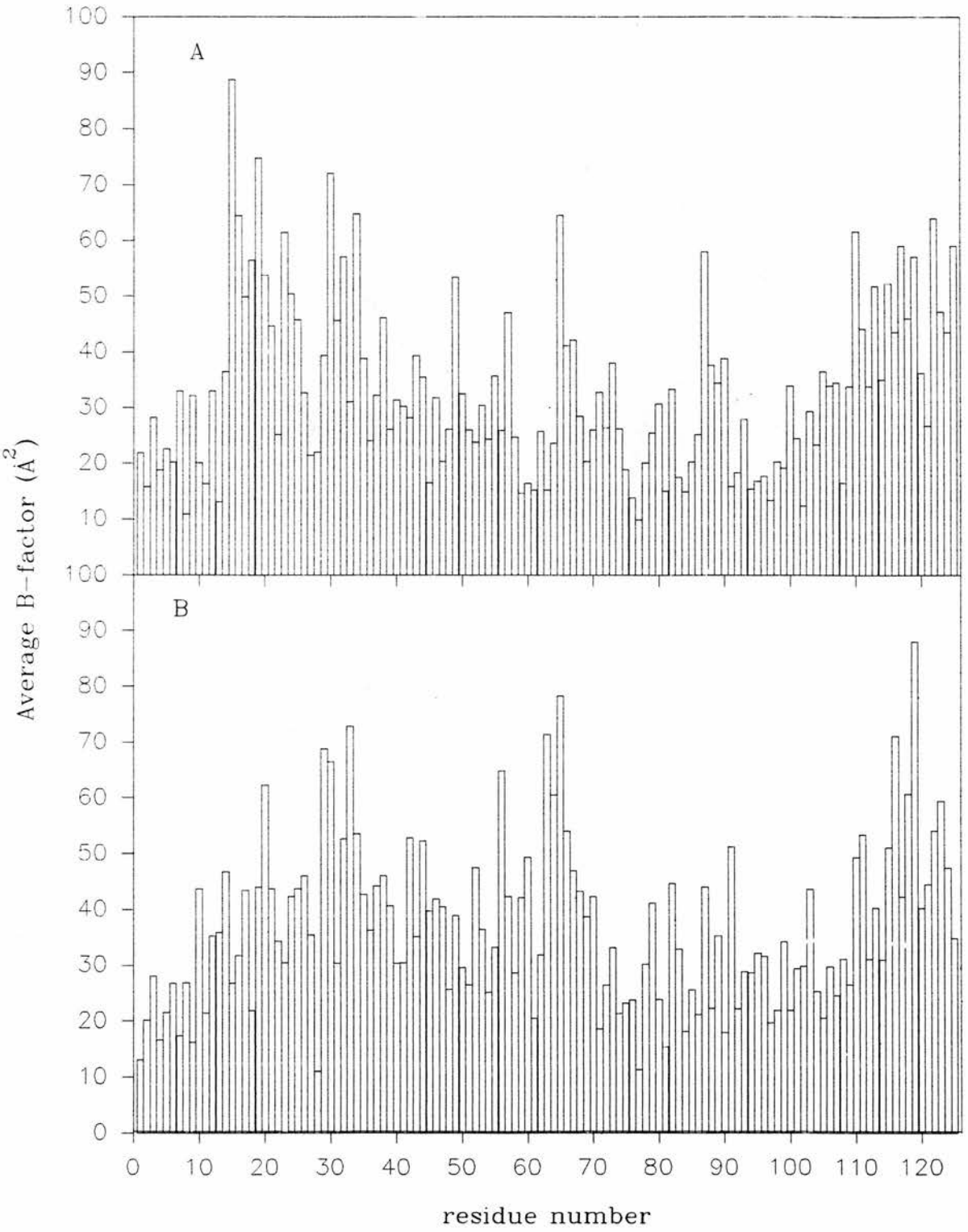
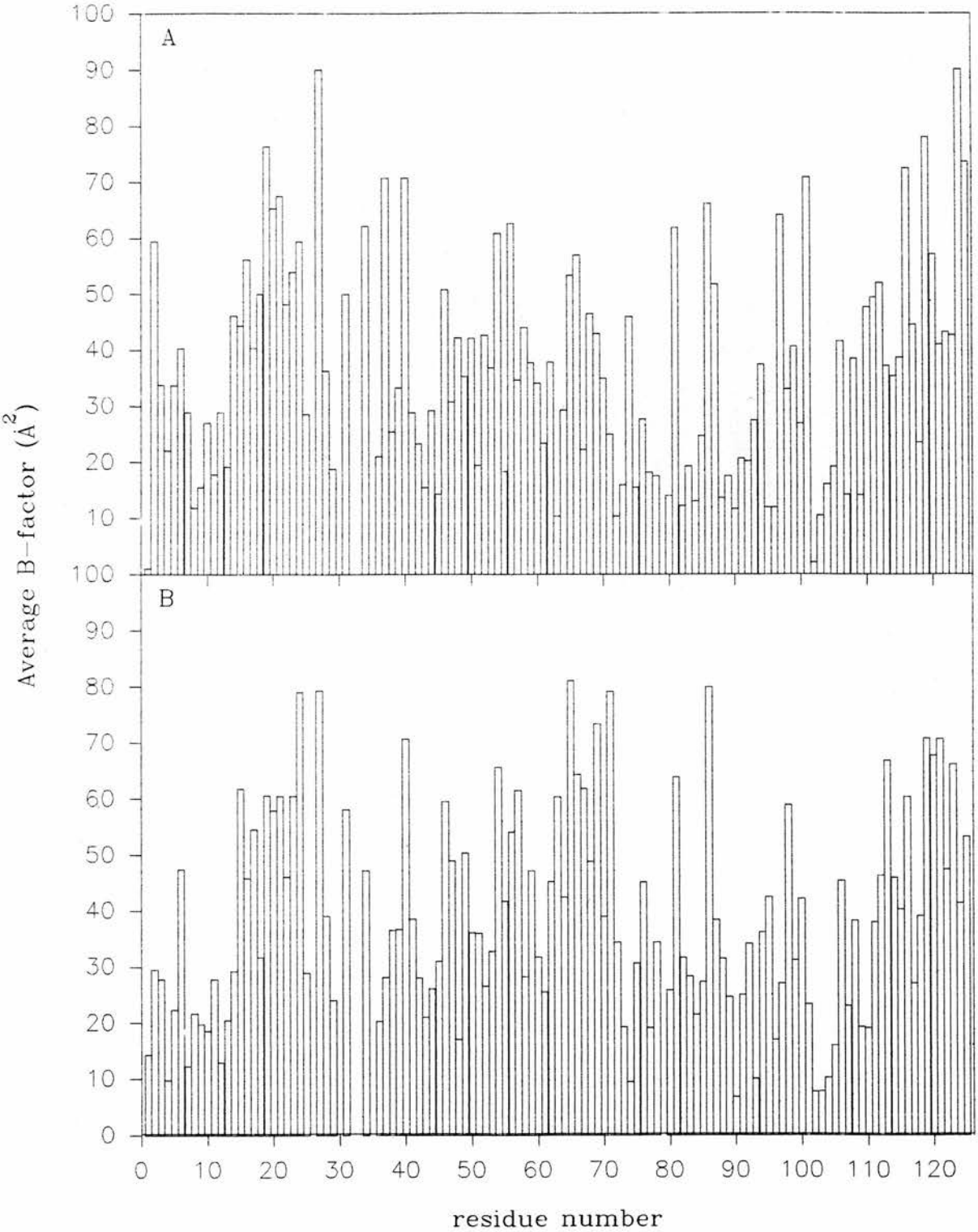


Figure 5.10 average B-factor for sidechain atoms of final model from TNT. Molecule A and molecule B are shown.



these badly defined loop regions the quality of the electron density map is good.

5.4 Refinement of putative PLA-2 - inhibitor complex

Crystals of the proposed complex were grown as outlined in Chapter 4. The data set obtained from a crystal of the proposed enzyme inhibitor complex from 15.25 - 2.07 Å contained 18 262 reflections of which 8669 were unique. The unit cell information of the proposed PLA2-drug complex is shown in table 5.6 together with the differences which relate this unit cell to that of the native unit cell. Attempts were made to scale this data with that used in the refinement of the native structure using the CCP4 program ANISOSC. The two data sets did not scale together well (R_{scale}=30% using F^2) and consequently the data used in the refinement was unscaled to the native. It is possible that the difference in the unit cell could account for the poor scaling. Although these two cells are similar, a difference map with coefficients F_O-F_C using the native structure in its unit cell with the data for the proposed complex did not show any signs of continuous density either in the active site or elsewhere in the asymmetric unit. It was possible that this lack of success was due to the differences in the unit cell dimensions. It was also possible that the two molecules which formed the asymmetric unit of the native structure may have been in different orientations relative to each other and the cell axes in the proposed complex. It was therefore necessary to perform a rigid-body refinement of the model within the unit cell of the complex. This was initially attempted using TNT by treating both molecules as separate independent rigid groups using the COMBINE option in the control program. Positional parameters are treated as

Table 5.6
Unit cell information for proposed PLA-2-drug complex. Also shown are the differences between the proposed PLA-2-drug unit cell and the unit cell of the native. Unit cell lengths are given in Angstroms, angles in degrees.

	a	b	c	α	β	γ
Proposed PlA2-drug complex	84.53	45.83	65.39	90.0	102.62	90.0
Native	83.20	45.44	65.13	90.0	101.98	90.0
Difference (Complex-native)	1.33	0.39	0.26	0	0.64	0
% Difference	1.6	0.85	0.4	0	0.6	0

though the groups are a rigid body. The control program fits an overall rotation and translation to the individual shifts of the atoms by a least squares fitting procedure. This was performed initially using all the data from 15 - 8 Å with fixed occupancies and temperature factors. The model used in this refinement contained no water molecules. Ten cycles of CG refinement were performed and the final structure analysed. No overall improvement in the R-factor was achieved and the rms movement after the tenth cycle was 0.001 Å. Analysis of the mean movements of the rigid group indicated that there was no net movement, but that both the rigid groups were making very small movements, like vibrations or oscillations, where they would be moved in one cycle and then be moved back again the next. This was not improved when higher resolution data from 15 - 2.4 Å was used.

It was then decided to try a different approach using XPLOR. Rather than perform a rigid body refinement using XPLOR, the native structure was used with the new cell parameters and the data of the drug complex in a simulated annealing refinement. An ideal weight W_A was calculated and the complete refinement was carried out as outlined in figure 5.11 for the data between 10 - 3.0 Å. Both a $2F_O$-F_C and an F_O-F_C were calculated for this structure and inspected using FRODO. While the density around the structure was quite good there was no continuous density in the F_O-F_C map, even when contoured at 0.5σ, in the active site of both molecules. The majority of the density was around the outside of the molecules in the asymmetric unit, with very little density within the molecule itself. Attempts were made to locate the bromine atom of the inhibitor by contouring at 2 and 3σ, but the positions of these electron dense regions bore no relation to the active sites of either molecules.

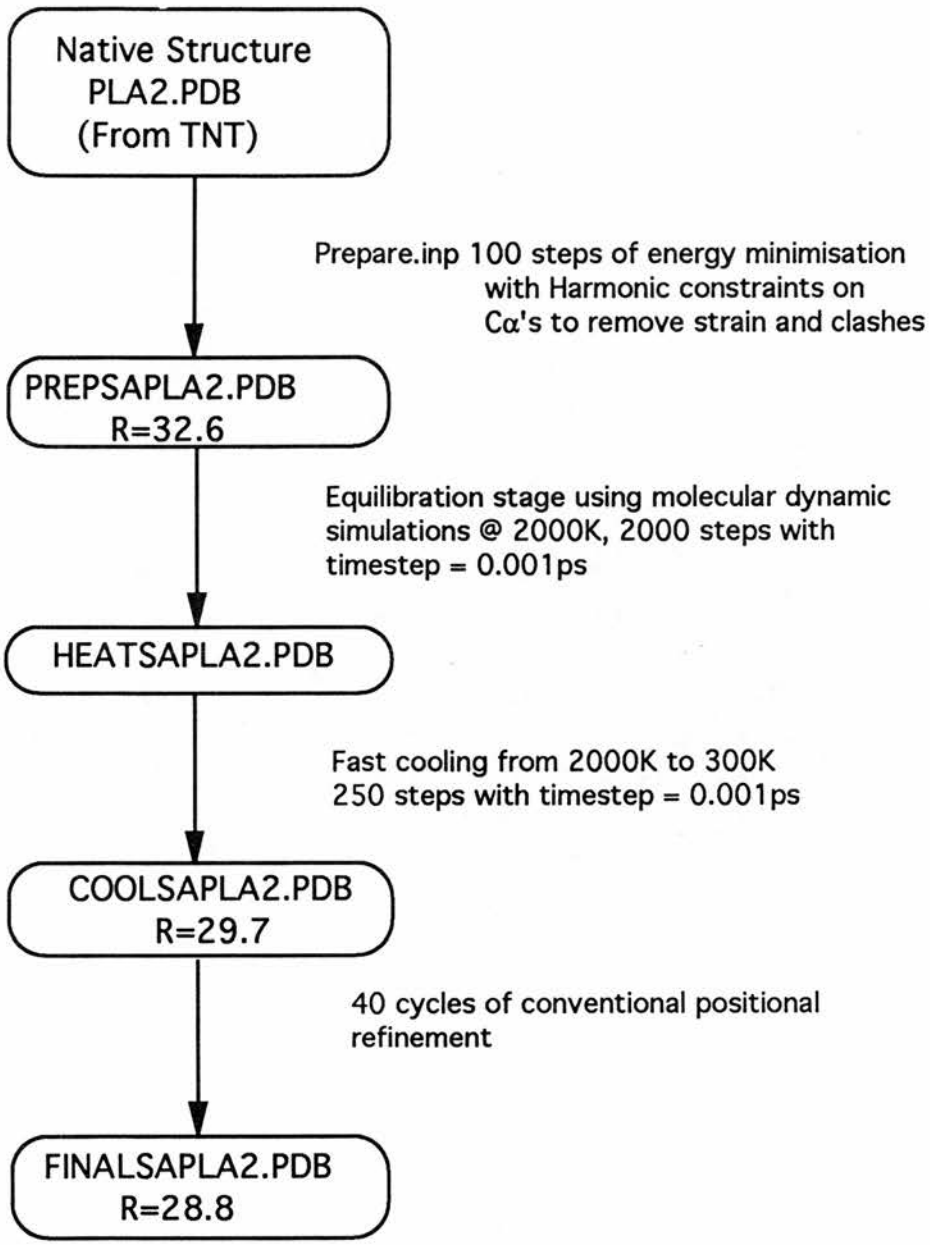


Figure 5.11 Annealing protocol from XPLOR for refinement of proposed drug-PLA-2 complex.

5.4.1 Discussion

The initial signs that a complex of enzyme and inhibitor had been crystallised - a slightly different crystal morphology and a slightly bigger unit cell, unfortunately have not stood up to further investigation. No traces of density were observed in the active sites of either molecule in the asymmetric unit or elsewhere in the crystal structure. Since no density could be assigned to the inhibitor molecule it must be assumed that there was no inhibitor bound at the active site or at any other location in the crystal. Consequently during the crystallisation experiments the hydrophobic inhibitor must have precipitated. Phospholipid like inhibitors such as the one used here are not sufficiently tight binding to be regarded as binding irreversibly. It is probable that on release from the active site during the crystallisation trial the inhibitor precipitated. This is further substantiated by the inability to locate the relatively electron dense bromine atom which forms part of the inhibitor molecule. Thus it would appear that the crystal structure from this data is very similar to the TNT structure refined from the so called "native" data despite the unit cell differences. Superposition of the C_α's of the TNT structure onto this structure (rms deviation for the two molecules of 0.419 Å) supports this idea.

Chapter 6 The crystallographic structure of equine phospholipase A2

Chapter 6

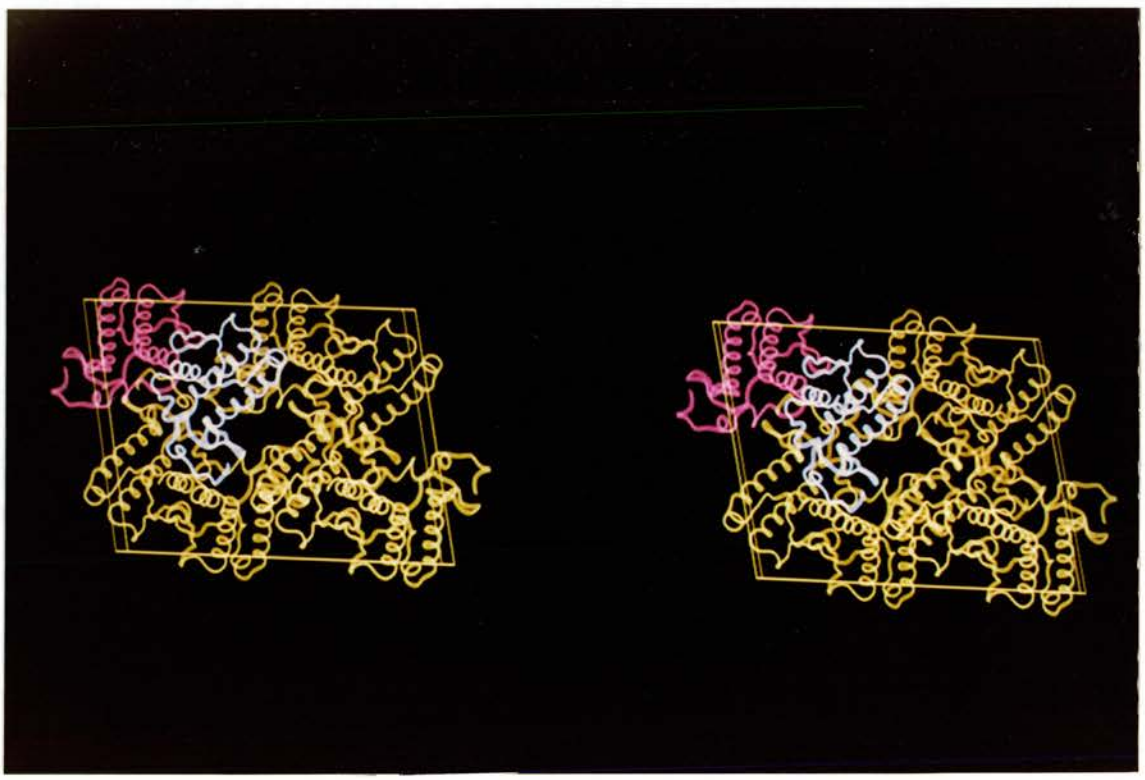
6.1 Introduction

The purpose of this chapter is to analyse the structure of equine phospholipase A2 (PLA-2) which has been refined from x-ray data as outlined in Chapter 5. The structure is interpreted in relation to its biological activity and compared and contrasted to the structure of bovine PLA-2 (Dijkstra *et al.*, 1981).

6.1.1 Symmetry

Equine PLA-2 crystallises in the monoclinic space group C2. The asymmetric unit is a dimer and there are four symmetry related dimers in the unit cell. Plate 6.1 shows a stereo view down the b-axis. The two molecules A and B are coloured pink and white respectively, the other molecules, related by symmetry, are coloured as yellow dimers.

Plate 6.1 Stereo picture of the unit cell of equine PLA-2.



6.2 Folding of the polypeptide chain and secondary structure.

From the ribbon structure shown in plate 6.2 three long helices A, C and E (using the same notation as for the bovine enzyme), and two shorter ones B and D can be seen. Helix A runs from residue 1 to 13. Helix C is 19 residues long and begins at residue 40. Helix E is of a similar length (20 residues, from 89 to 108). Helix B (residues 19 to 23) and helix D (residues 59 to 64) are both short α-helices of about one turn. There are two antiparallel ß-strands, running from residues 74 to 78 and 81 to 85 respectively. The structure was analysed for the occurrence of ß-turns using the DSSP program of Kabsch and Sander (1983). The criteria adopted for the definition of the turn were from Chou and Fasman (1977). Table 6.1 lists the ß-turns in both molecules of equine PLA-2.

Since the equine sequence is 78.6% identical with that of the bovine enzyme it is not surprising that the structure of equine phospholipase is very similar to that of the bovine enzyme. Plate 6.3 shows a stereo picture of the backbones of the two molecules A and B, after optimal superposition to the backbone atoms (1 to 121) of the bovine enzyme using SYBYL. A comparison of the C_α's in molecule A with those of the bovine enzyme gives a root mean square difference of 0.951 Å, similarly for molecule B this is 0.868 Å. We have already seen that this rms distance between A and B is 0.789 Å.

There were four major areas where there appeared to be a significant difference to the bovine structure. This was the case for both molecule A and B. It was also noticed that these four areas coincided with the areas of difference between molecule A and B. All of these areas were to be found in certain loop regions and not in any of the secondary structural elements. On analysis, it was seen that three out

Plate 6.2 Stereo picture of the ribbon structure of equine PLA-2. The labels illustrate the limits of the secondary structural elements.

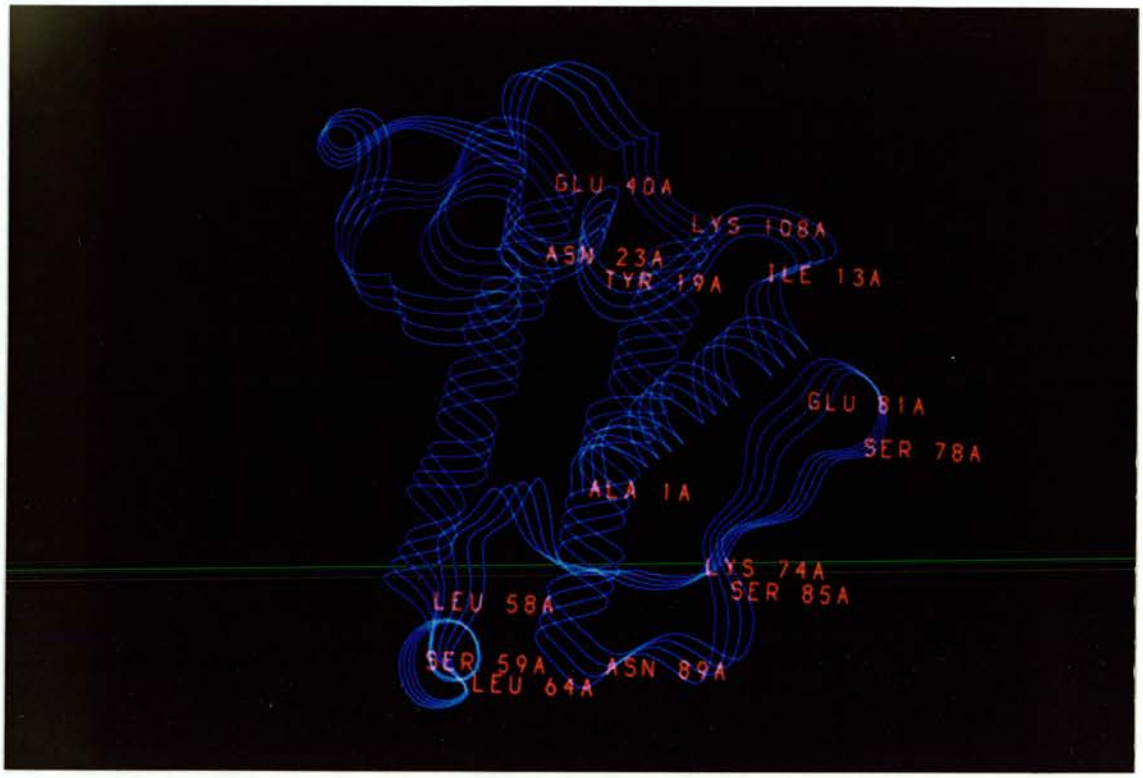
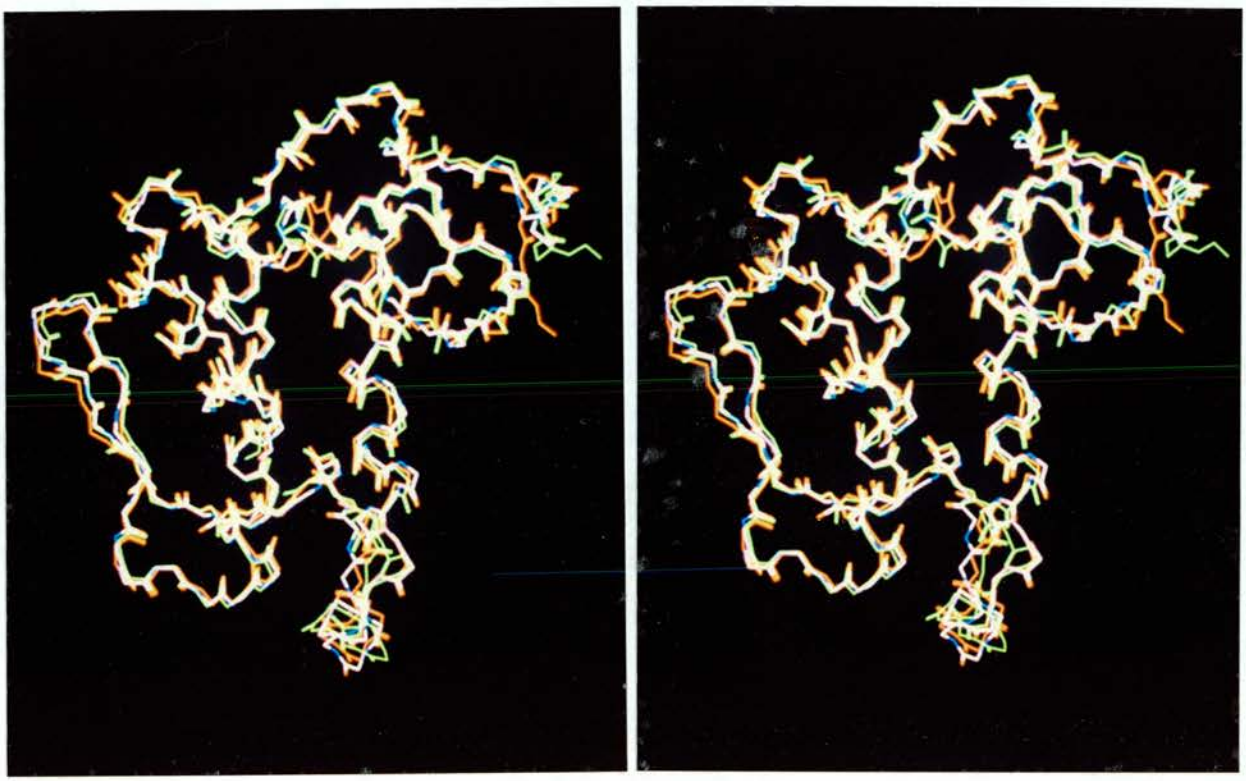


Table 6.1 β - turns in equine PLA-2.

Segment	Residues	ϕ_2	ψ_2	ϕ_3	ψ_3	Type
55A-58A	Ala.Lys.Glu.Leu	-51.7	-21.0	-64.2	-34.8	III
67A-70A	Asn.Pro.Tyr.Thr	-36.4	-27.5	-85.6	-7.7	I
78A-81A	Ser.Gly.Thr.Glu	79.5	-119.5	-84.2	-17.5	II'
85A-88A	Ser.Asp.Lys.Asn	-49.7	-39.0	-68.2	-18.0	III
113A-116A	Pro.Glu.Asn.Lys	-68.5	-34.9	-78.3	-33.4	III
13B-16B	Ile.Pro.Asn.Ser	-63.2	-24.5	-96.8	7.2	I
58B-61B	Leu.Ser.Ser.Cys	-53.1	-46.1	-69.9	-22.9	III
78B-81B	Ser.Gly.Thr.Glu	61.7	-129.6	-99.7	10.7	II'
85B-88B	Ser.Asp.Lys.Asn	-67.3	-15.7	-66.2	-18.3	III
112B-115B	Asn.Pro.Glu.Asn	-65.4	-17.5	-58.8	-37.9	III

Plate 6.3 Stereo picture of the backbone fit of equine PLA-2 with bovine PLA-2. Molecule A is shown in red, molecule B in green. The bovine backbone is coloured by atom type. The superposition and picture were produced from SYBYL.



of the four could be explained by different crystal packing interactions.

The first region of difference was between residues 15 and 22. It was observed that in molecule A the Asn 15 sidechain formed a strong hydrogen bond to the sidechain of Arg 122 of molecule B related to it by the symmetry operator ½-x, y-½, 1-z. The sidechain of Asn 15A also formed a second hydrogen bond to the sidechain of Asn 115A. In molecule B there was no such hydrogen bonding of Asn 15B since there were no packing interactions in this region. Additionally Tyr 19B OH formed a strong hydrogen bond with Glu 114B related by the symmetry operator ½-x,½+y,1-z. The second region of difference was between residues 56 and 65. Here three hydrogen bonds were formed between molecule A and molecule B of two independent symmetry related molecules. These were between the sidechains of Arg 62A and Glu 89B, Asn 67A and the mainchain carbonyl of Ser 59B related by the symmetry operator ½-x,½+y,1-Z the mainchain carbonyl of Phe 63A and the sidechain of Lys 74B related by the symmetry operator x-½,½+y,z. These hydrogen bonds were not possible in molecule B and there were no crystal contacts for this area. The third region of difference was between residues 114 and 120. In molecule A it was noted that the sidechain of Glu 114A formed a strong hydrogen bond to the sidechain of Asn 46B related to it by the symmetry operator x-½,½+y,1-z. Similarly the sidechain Gln 115A formed a strong hydrogen bond with Gln 112A. However in molecule B Glu 114B formed no hydrogen bonds to any symmetry related molecules and the sidechain of Gln 115B formed two hydrogen bonds. The first of which is to the sidechain of Asp 39B and the second is to the sidechain of Gln 15A related to it by the symmetry operator ½+x,½+y,z. This is the same hydrogen bond as

described for the first region. The fourth region of difference between 30 and 35 could not be explained by different crystal contacts. A possible explanation for the differences will be discussed in section 6.5.

6.3 The disulphide bridges.

Plate 6.4 shows the location of the seven disulphide bridges in the equine structure. These clearly help to stabilise the enzyme structure and are responsible for its ability to survive the high temperatures involved in the purification steps outlined in Chapter 3. These disulphides are also important for the enzyme's biological activity by providing the necessary structural stability to work at the micelle interface. The geometry of the disulphides are shown in table 6.2. The X_3 angle is defined according to the IUPAC-IUB (1970) convention. In agreement with the findings of Thornton (1981), the X_3 angles are distributed around $X_3 = +90^\circ$ (right handed) and $X_3 = -90^\circ$ (left handed). Similarly local disulphides (half-cystine separation $<$ 45 residues) were right handed (6 out of 6). For non-local disulphides (half-cystine separation $>$ 45 residues), these were equally likely to be left handed (4 out of 8). The geometry of the disulphide bond 27 - 124 is particularly variable and this is possibly due to the differences in the main chain conformation at the C-terminal region between molecule A and B and the mobility of this region. Analysis of the thermal factors of both the main and sidechains of residues 120 - 124 (figure 5.9 and 5.10) and the different conformations found between A and B would indicate that these five residues at the C-terminus are mobile. The mobility of these residues has been noted in the 2.1 Å structure of an engineered porcine PLA-2 (Thunissen *et al.*, 1990),

Plate 6.4 Stereo picture of the ribbon structure of equine PLA-2 showing the location of the disulphide bridges.

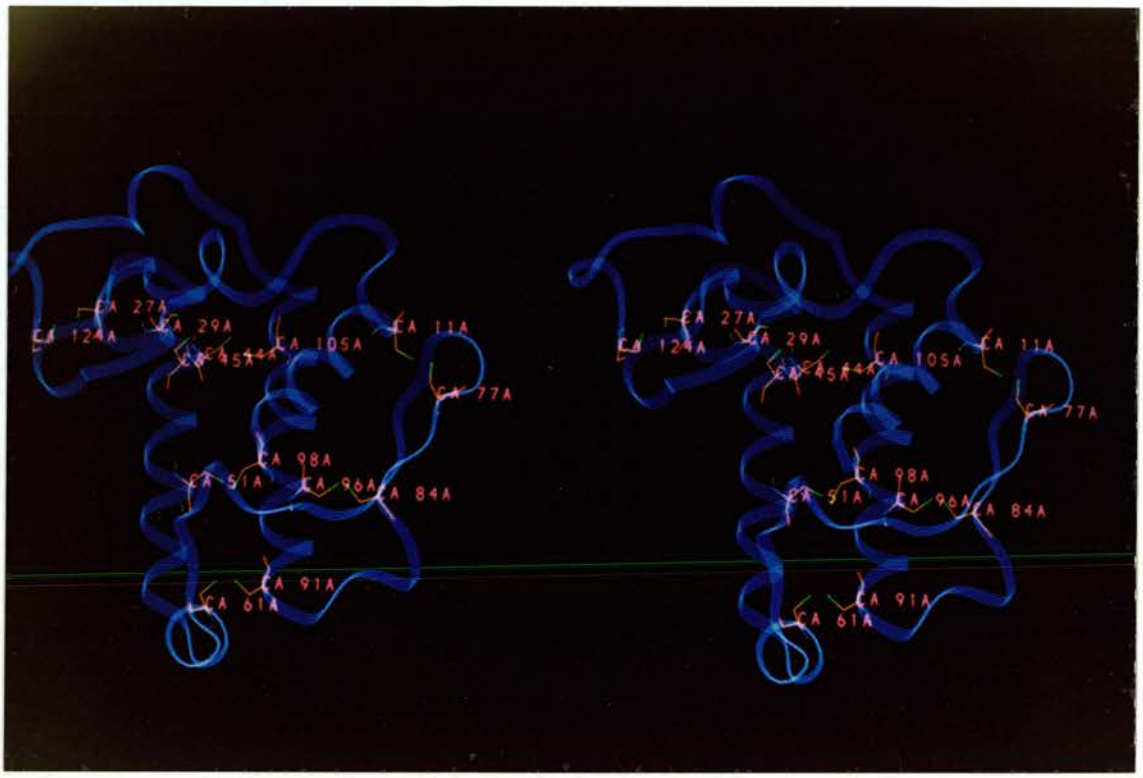


Table 6.2 Torsion angles (χ_3) of the disulphide bridges in equine PLA-2.

Location	Torsion Angle (χ_3)*			
	molecule A	Hand	molecule B	Hand
Cys 29 - Cys 45	92	R	92	R
Cys 44 - Cys 105	-82	L	-88	L
Cys 51 - Cys 98	-83	L	112	R
Cys 61 - Cys 91	100	R	98	R
Cys 77 - Cys 11	94	R	92	R
Cys 84 - Cys 96	89	R	93	R
Cys 124 - Cys 27	-77	L	110	R

*χ_3 angles are defined according to IUPAC-IUB (1970) convention.

where the two molecules that formed the asymmetric unit also possessed differing conformations in this region.

6.4 The dimer interface.

The interaction between the two molecules of the asymmetric unit occurs through a number of hydrogen bonds. Two secondary structural elements of the enzyme are involved. Plate 6.5 illustrates the nature of the dimer interface and also how the two molecules are positioned relative to each other in the unit cell. There is a number of hydrogen bonding interactions which involve residues from the helix A (Ala 1 - Ile 13) of both molecules. These helices are adjacent and in an antiparallel arrangement. As a consequence of this there are four hydrogen bonds formed between the two chains as listed in table 6.3. Two of these occur in the middle of the helix between Ser 7 A and B. The third and fourth hydrogen bonds are identical with respect to the donor and acceptor groups involved but have slightly different geometries. At both ends of the helices the sidechain NE1 of Trp 3 hydrogen bonds to the mainchain carbonyl of Gly 79 of the other molecule. Gly 79 is at the middle of the ß-turn of the antiparallel ß-sheet. Thus the helix of molecule A is not only hydrogen bonded to the helix of molecule B but also the ß-sheet structure of molecule B.

The second structural element involved in the dimer interface is the ß-sheet. The ß-sheets of both molecules are adjacent along one edge and run in an antiparallel manner forming an extensive hydrogen bonding network as shown in plate 6.6. The hydrogen bonds and their geometries are listed in table 6.4. Additionally these sheet structures have an amphipathic distribution of amino acids. The hydrophilic residues Lys 74, Ser 76, Ser 78, Glu 81 and Thr 83 lie

Plate 6.5 Stereo picture showing the dimer interface and location of the two molecules in the asymmetric unit. Molecule A is shown in blue, molecule B in red and the unit cell in purple.

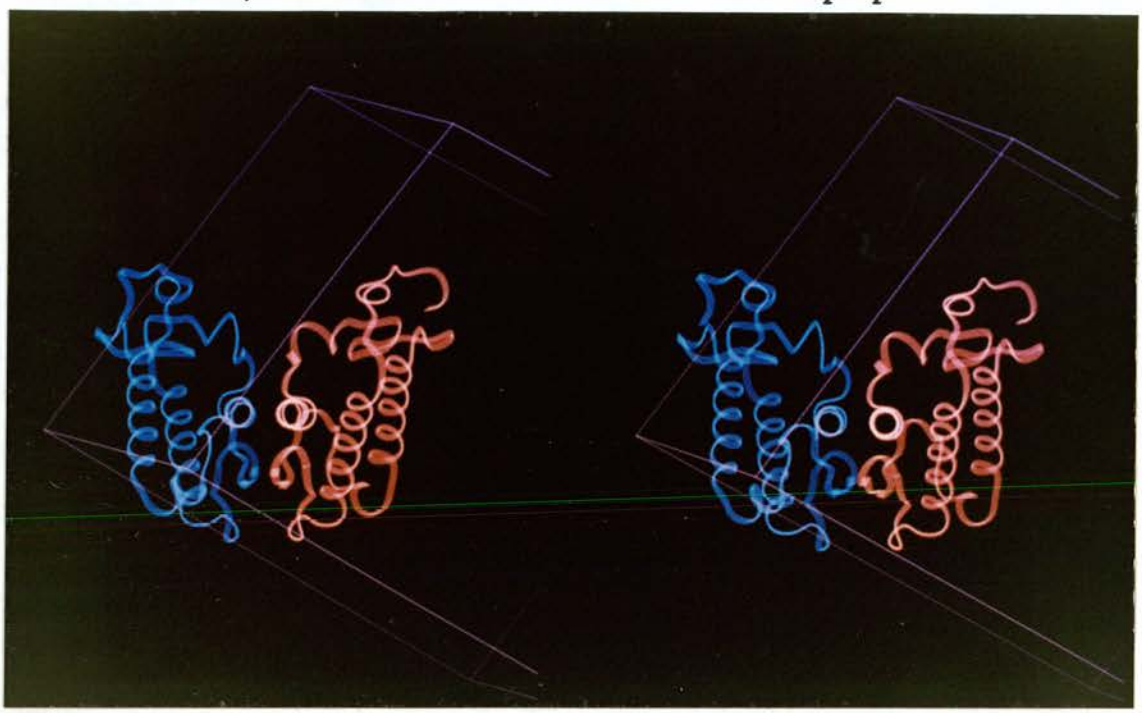


Plate 6.6

The antiparallel β-sheet/sheet interaction at the dimer interface. Both the ribbon pattern and the atoms are shown. Molecule A is shown in white, molecule B in blue. The hydrogen bonds are illustrated by the broken lines. In this orientation the hydrophobic residues are behind the plane of the sheet. The electron density map (2Fo-Fc) was contoured at 1σ.

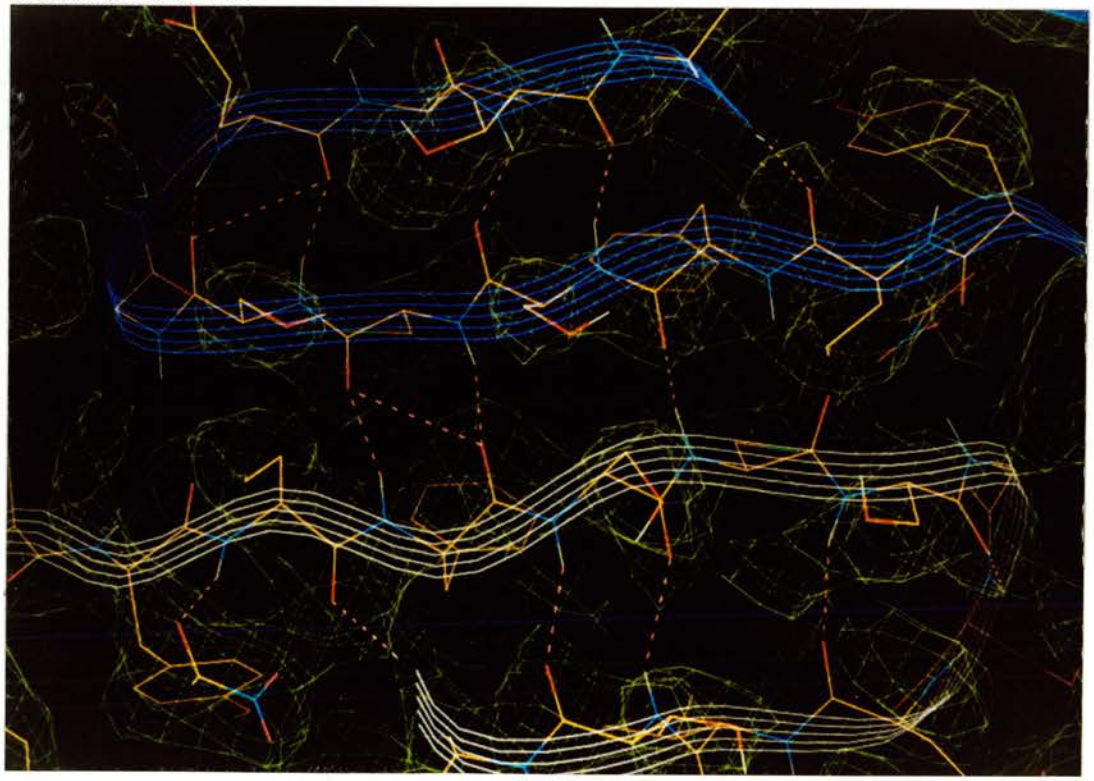


Table 6.3 Hydrogen bonding geometries of the groups involved in the helix-helix interaction of the equine PLA-2 asymmetric unit.

Description	Distance (Å)		Angle (°)
OG 7B---OC 7A	3.2	(O-O)	130
OG 7A---OG7B	2.7	(O-O)	116
NE1 3B---OC 79A	2.6	(N-O)	171
NE1 3A---OC 79B	3.4	(N-O)	107

Table 6.4 Hydrogen bonding geometries of the groups involved in the sheet-sheet interaction of the equine PLA-2 asymmetric unit.

Description	Distance (Å)	Angle (°)
NH 75A---OC 77B	3.0 (N-O)	162
NH 77B--OC 75A	2.8 (N-O)	148
NH 77A--OC 75B	2.9 (N-O)	130

above the sheet exposed to solvent. The hydrophobic residues Phe 75, Cys 77, Val 82 and Cys 84 lie below the sheet structure. The hydrophobic residues below the sheet provide additional van der Waals contacts. The amphipathic partitioning of these residues must also contribute further to the stability of the dimer. As a combination of these three interactions, the dimer of the asymmetric unit is tightly associated.

6.5 The active site.

The active site of the equine enzyme is located in a cavity at the molecular surface with His 48 located at the bottom. The imidazole ring of His 48 forms a hydrogen bond with the carboxyl group of Asp 99, similar to the serine protease enzymes. The wall of the active site is covered by hydrophobic residues: Phe 5, Ile 9, Phe 22, Ala 102 and 103, Phe 106 and the disulphide bond between Cys 29 and 45. Plate 6.7 shows a stereo picture of the position of the active site, including the calcium ion, the liganding oxygens, and the His 48 - Asp 99 couple. The refinement of equine PLA-2, as described in Chapter 5, was completed with only partial occupancies for both calcium ions. As already mentioned the quality of the electron density of both the loop itself and that of the calcium ion merited this. In particular the electron density of residues 30 and 31 was poor and did not improve when the calcium ions were given partial occupancies. It has been noted that Leu 31, can adopt different conformations (Dijkstra *et al.*, 1983, Thunissen *et al.*, 1990) which has been attributed to conformational freedom in this region. The geometry of the oxygen ligands to the active site calcium ion are shown in table 6.5 a-b. The different geometries could not be attributed to sequence variations

Plate 6.7 Stereo picture of the Cα trace of equine PLA-2 showing the location of the calcium ion of the active site. Also shown are the mainchain atoms from 28-32 which form the calcium binding loop.The His 48 - Asp 99 couple of the active site is also shown. The two main helices C and E are running vertically

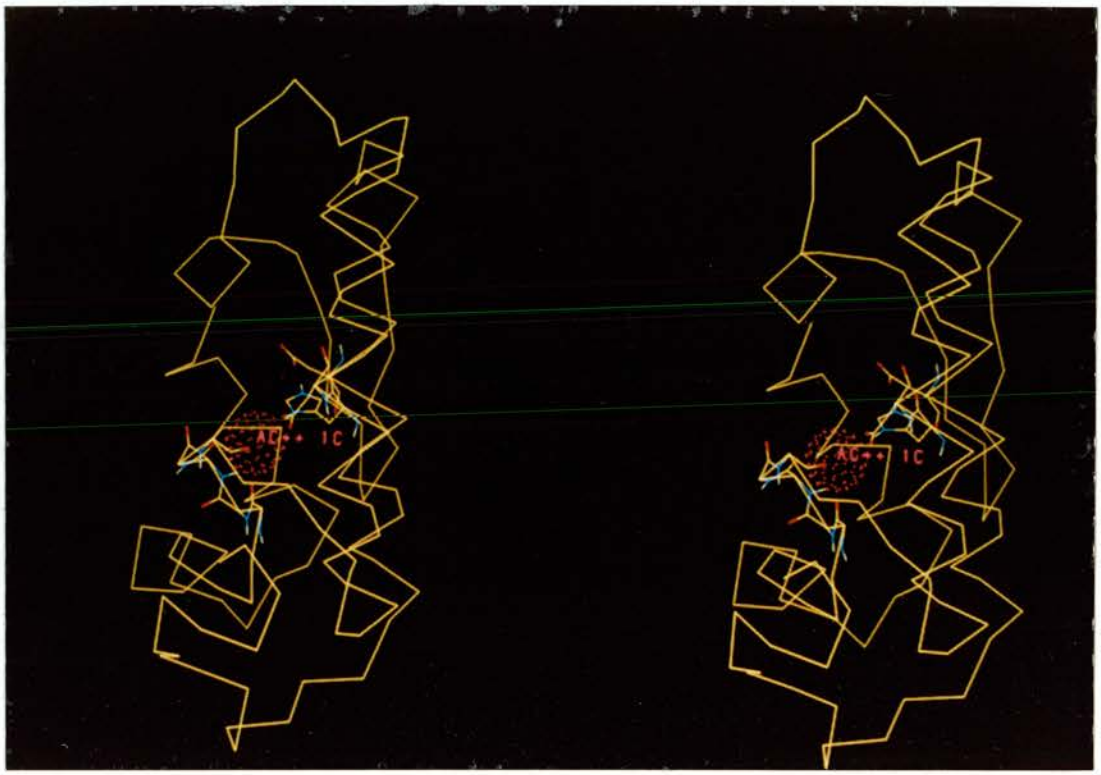


Table 6.5a Ligand to calcium distances in the active site of equine PLA-2 and a comparison to bovine PLA-2.

Calcium ligand	Distance (Å)		
	Molecule A	Molecule B	Bovine *
OD1 49	1.90	2.40	2.48
OD2 49	2.39	2.88	2.69
O32	2.55	2.28	2.30
O30	5.46	3.08	2.47
O28	3.23	3.09	2.29

*from Dijkstra et al., (1981).

Table 6.5b Ligand to calcium angles in the active site of equine PLA-2 and a comparison to bovine PLA-2.

Description	Angle (°)		
	Molecule A	Molecule B	Bovine *
OD1 49 - Ca++ - OC 32	111	92	88
OD1 49 - Ca++ - OC 30	147	148	164
OD1 49 - Ca++ - OC 28	134	116	97
OD2 49 - Ca++ - OC 32	158	134	135
OD2 49 - Ca++ - OC 30	92	100	141
OD2 49 - Ca++ - OC 28	96	98	109

* from Dijkstra et al., (1981).

between the enzymes, since the active site residues are highly conserved. While it is clear that the geometry is similar there is a significant discrepancy in some of the liganding distances. This is possibly as a consequence of disorder caused by the partial occupancy of the calcium ion. Disordered atoms (or atoms with partially occupied sites) may be separated by distances less than the minimum hydrogen bond length (Karle & Duesler, 1977). The same phenomenon has been observed for the ligand to calcium distances in trypsinogen, where the calcium site was only half occupied and the average calcium to ligand distance was 2.1 Å instead of the expected 2.4 Å (Fehlhammer *et al.*, 1977). Thus the rather short liganding distances of the Asp 49 carboxylate group can be explained by the partial occupancies of the calcium ion. The geometry of the carbonyl of Gly 30 is considerably different from that of the bovine enzyme. No water ligands for the calcium ions were found around the calcium of either molecule although a water molecule was located in the active site of each molecule. Each formed a single hydrogen bond to an active site sidechain as indicated in table 6.6. It is clear these are not similar to those identified in the bovine structure and are not in the correct geometry for the catalytic mechanism.

The other residues of the calcium binding loop were well defined and the loop was seen to be stabilised by a number of interactions with other residues or mainchain atoms as listed in table 6.7. It is noticeable that residues Tyr 25, Tyr 28, Asp39 and Asp42 which are involved in stabilizing the calcium binding loop are invariant in all the known pancreatic and snake venom phospholipases (Heinrikson *et al.*, 1977). The loop is further stabilised by the disulphide linking Cys 27 to Cys 124 and the disulphide bond between Cys 29 and Cys 45.

Table 6.6 Geometry and hydrogen bonds of active site waters in equine PLA-2 and their comparison to bovine PLA-2 (Dijkstra et al., 1981).

Hydrogen bond distance (Å)		Angle measured	Angle (°)	
				Bovine
48A ND1--OH2 4D	2.2	O32A, Ca++, OH2 4D	140	150
		OD1 49A, Ca++, OH2 4D	96	75
49B OD2--OH2 5D	2.8	OD1 49B, Ca++, OH2 5D	72	76
		OD2 49B, Ca++, OH2 5D	24	94

Table 6.7 Interactions which stabilise the calcium binding loop.

Description	Distance (Å) Molecule A	Molecule B
NH 28 -- OD1 42	2.8	2.9
NH 24 -- OC 29	2.7	3.2
NH 29 -- OC 25	3.4	2.9
OH 28 -- OC 35	2.3	2.3
NH 27 -- OD1 42	2.6	2.8
NH 38 -- OD 42	3.4	3.0
SG 27 -- SG 124	2.0	2.1

It can be concluded that these invariant residues have the important function of stabilising the calcium binding loop and, together with Asp 49, they guarantee a precise three dimensional geometry of the calcium ion with respect to the active site residues. Thus the majority of the loop, in the absence of the calcium ion, does not have the ability to adopt multiple or different conformations. However, analysis of the thermal factors of the residues in the active site (figure 5.9 and 5.10) would also indicate that there was mobility in the part of the calcium loop and that this mobility was due to the partial occupancy of the calcium ion. The porcine active site (Dijkstra *et al.*, 1983) was found to be almost identical to that of the bovine enzyme. Additionally the active site residues possessed thermal factors that were 60% lower than the average temperature factor for all the protein atoms which would indicate that the bound ion heavily reduces the mobility of the active site residues.

6.6 The amino terminus.

The N-terminal region and, in particular, the α-NH_3+ group of Ala 1 play an important role in the activity of the enzyme on aggregated substrates (Pieterson *et al.*, 1974a; van Dam-Mieras *et al.*, 1975). Ala 1 has its amino group buried in the interior of the enzyme, and this amino group has various hydrogen bonding interactions. In the bovine and porcine enzymes, a similar hydrogen bonding network is present around an internal water, but they differ in their finer details as shown in figure 6.1. In the porcine enzyme, the N-terminus is somewhat more open to the solvent than the bovine enzyme (Dijkstra *et al.*, 1983). The N-terminal region of equine phospholipase is depicted in plate 6.8. Plate 6.9 shows the quality of the electron density around

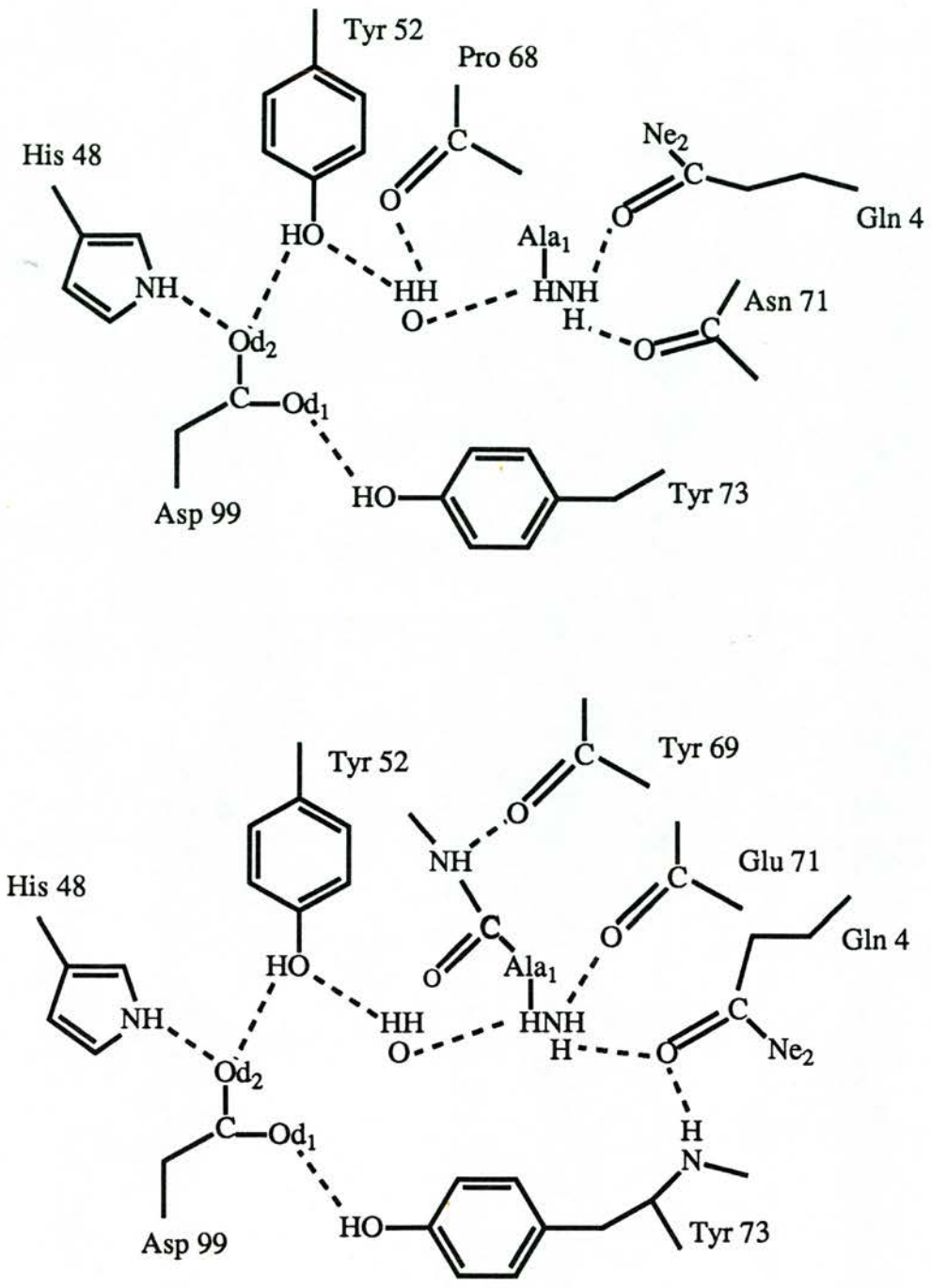


Figure 6.1 Hydrogen bonding network around the amino terminus of bovine (upper) and porcine (lower) phospholipase A2. Adapted from Dijkstra et al., (1983).

Plate 6.8 Stereo picture of the amino-terminal region of equine PLA-2. The hydrogen bonding network in the amino terminal region of molecule B is illustrated by the broken lines.

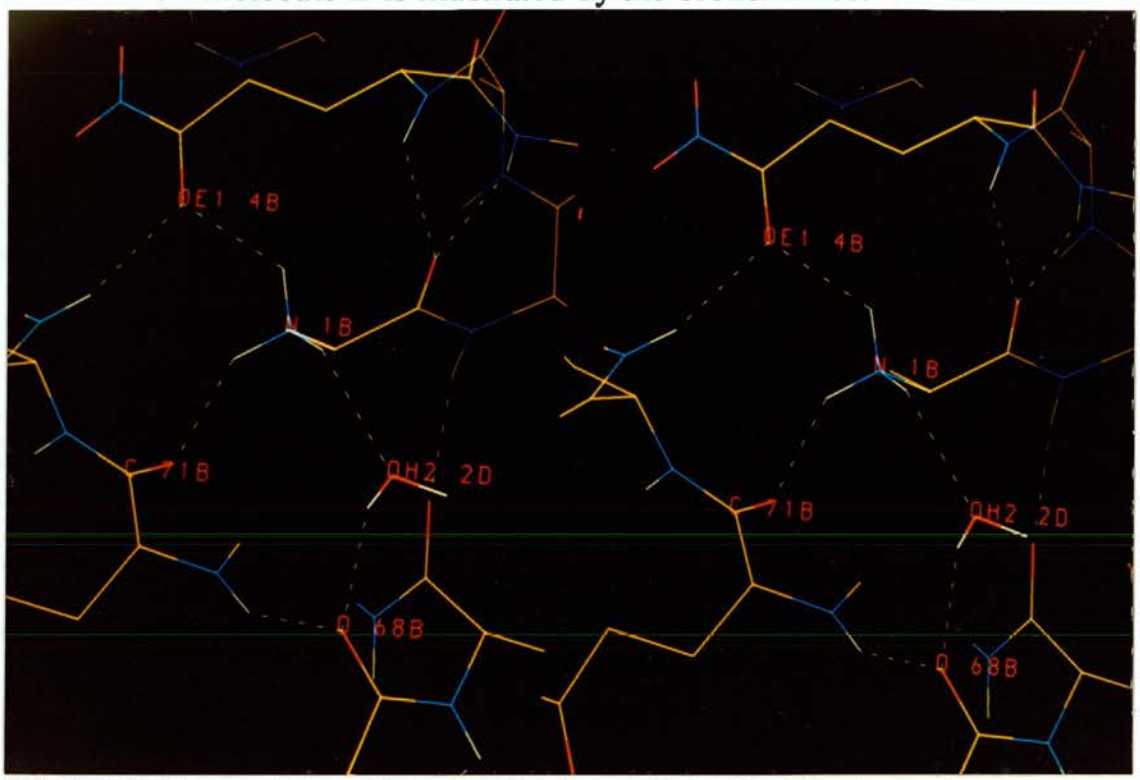


Plate 6.9 Stereo picture showing the integrity of the electron density around the water at the amino terminus in equine PLA-2. This is shown for molecule A and the electron density map (2Fo-Fc) has been contoured at 1.0σ.

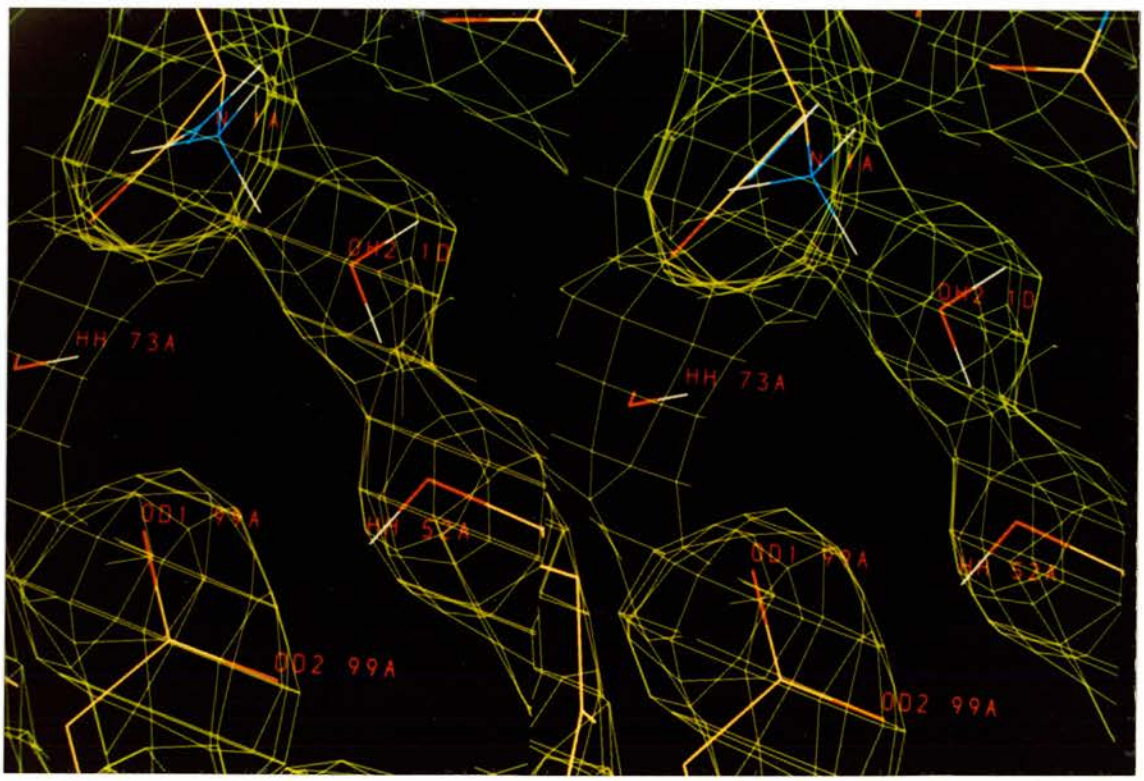


Table 6.8 Hydrogen bonding distances in the N-terminal region.

From	To	Distance (Å)	
		Molecule A	Molecule B
Ala1 N	4 OE1	-	2.8
Ala1 N	71 O	2.7	2.9
Ala1 N	Water	2.3	3.1
Water	52 OH	2.6	3.1
Water	68 O	3.1	3.2
Tyr52 OH	99 OD2	2.8	3.0
His48 NE2	99 OD2	3.1	2.6
Tyr 73 OH	99 OD1	2.6	3.1

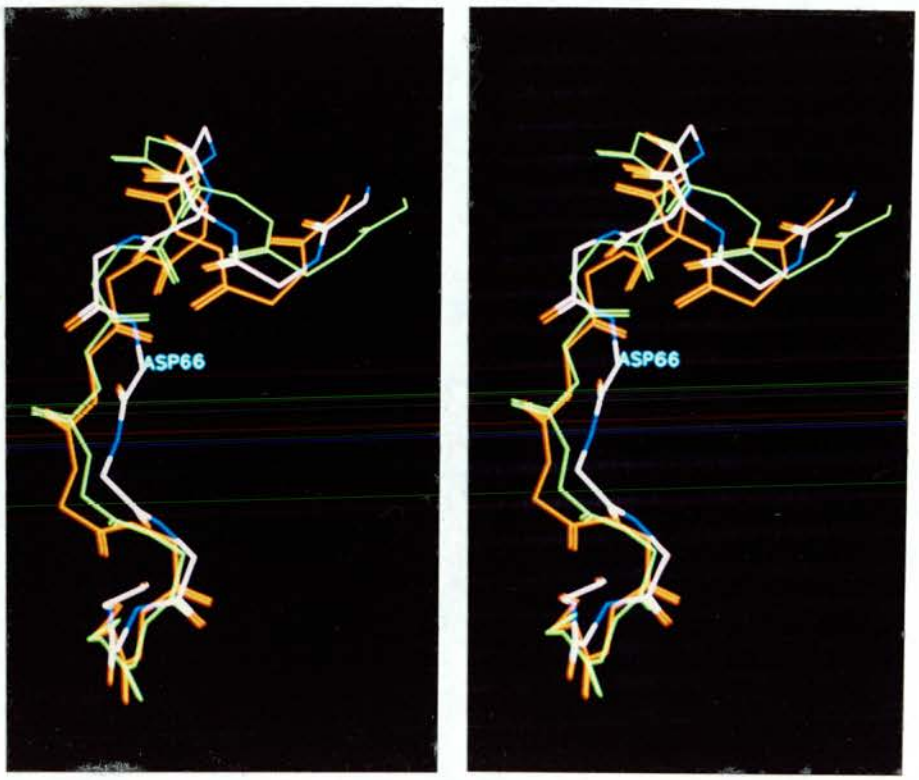
- no hydrogen bond was found.

the water molecule which is bound in this region. In table 6.8 the hydrogen bonds in the N-terminal region are listed. The hydrogen bonding network is very similar to that of the bovine enzyme as illustrated in figure 6.1. However there is a slight difference between both molecules of the asymmetric unit cell. In molecule A the sidechain of Gln 4 is not involved in hydrogen bonding to the NH3+ group, rather it is involved in hydrogen bonding to two mainchain atoms. This is not the case for molecule B, here the sidechain hydrogen bonds in the same manner as the bovine structure depicted in figure 6.1.

6.7 The '65' loop.

This is the name given to the loop region between Ser 60 and Thr 70. It has been suggested (Drenth *et al.*, 1987) that this region might be involved in the interaction of the enzyme with micelles. This loop region forms part of what has been called the interface recognition site (IRS). The IRS was initially described as a result of kinetic and amino acid modification experiments (Volwerk & de Haas, 1982) and this was subsequently completed when the X-ray structure became available. Plate 6.10 shows a stereo picture of the backbone atoms of A and B after optimal superposition of the bovine and equine structures. It is clear that while the conformation of the equine enzyme is slightly different, the loop has a similar overall fold to the bovine. Also this loop is slightly different between molecule A and B. The presence of different crystal contacts has been already given as a possible reason for this (Chapter 6.2). Further evidence that this loop is more like that of the bovine enzyme was seen in chapter 6.6 where the mainchain carbonyl of Pro 68 is involved in the hydrogen bonding

Plate 6.10 Stereo picture of the '65' loop backbone of equine PLA-2 after optimal superpositon with the bovine PLA-2. The region between 60-70 is shown for molecule A in green and molecule B in red, the bovine loop is coloured by atom type.



network at the amino terminus. As seen in figure 6.1 Pro 68 is not involved in hydrogen bonding at the amino terminus in the porcine enzyme. The fact that equine PLA-2 adopts the bovine conformation calls into question the explanation given as to why the porcine is different to the bovine (Dijkstra *et al.*, 1983). Here a single substitution in this loop region (bovine Val 63 -> porcine Phe 63) was suggested to be responsible for the differences in the conformation of the bovine and porcine structures. This was suggested to explain the different affinities of bovine and porcine PLA-2's for micelles (Dijkstra *et al.*, 1983). In the equine enzyme there is this similar single substitution but no dramatic conformational change is observed. However it is possible that crystal contacts are responsible for preventing the equine structure adopting the porcine conformation. It is more likely, however, that crystal packing interactions are responsible for producing the slight differences between the two molecules in the asymmetric unit and also to the bovine conformation.

6.8 The second calcium binding site.

In addition to the catalytically essential calcium ion, it has been demonstrated that porcine and equine PLA-2's contain a second calcium binding site. Compared with the calcium ion at the active site, the second ion binds with almost tenfold lower affinity. The importance of this site became clear from the observation that binding of a second ion is essential for effective interaction of PLA-2 with organised lipid/water interfaces, at alkaline pH (van Dam-Mieras *et al.*, 1975, Slotbloom *et al.*, 1978). Site directed mutagenesis experiments with porcine PLA-2 (van den Bergh *et al.*, 1989) have shown that two carboxylates, Glu 71 and, to a lesser extent, Asp 66, were both

directly involved in the low affinity calcium binding. However the crystal structure of porcine (Dijkstra *et al.*, 1983) has Asp 66 and Glu 71 too far apart to be both involved in ion binding. Athough in the bovine sequence residue 71 is Asn (which is presumably the reason it does not bind a second calcium), it is close to Asp 66 in the '65' loop of the bovine structure. Plate 6.11 shows this area for molecule A. It is clear that in the equine enzyme there are three carboxylates in close proximity, these are Asp 66, Glu 71 and Glu 92. Additionally there are a number of water residues around this site in both molecule A and B but no calcium could be identified at these sites. It is not surprising that calcium ions were not identified at these sites since the affinity of this second site is tenfold lower than that of the active site. Since the active site was only partially occupied the second site could not be expected to be filled, even partially.

6.9 Description of hydrogen bonding.

The hydrogen bond is an electrostatic interaction between a proton, carrying a partial positive charge, on a donor group, with the electron density of an acceptor atom. In proteins these normally involve amide groups, carbonyl oxygens, and water molecules. The limits used for the geometry of hydrogen bonds are those of Baker *et al.*, (1984). For example, the commonest hydrogen bond NH--O:

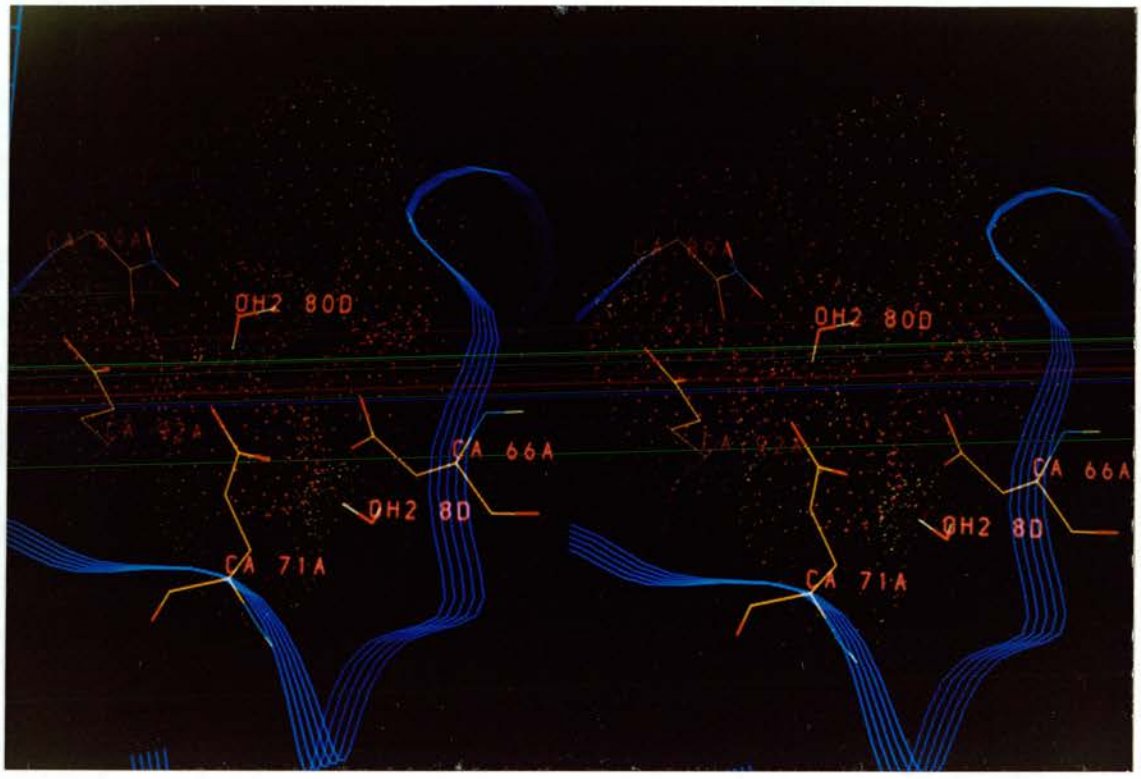
$$1.3\text{Å} \leq \text{O--H} \leq 2.5\text{Å}$$

$$90^{\circ} \leq \text{O--H--N} \leq 180^{\circ}$$

$$90^{\circ} \leq \text{C==O--H} \leq 180^{\circ}$$

However it must be noted that the relative strength of a particular hydrogen bond is highly dependent on it's precise geometry.

Plate 6.11 Stereo picture showing the proposed second calcium binding site of equine PLA-2. The enzyme is represented by the ribbon structure. Some potential liganding sidechains are shown along with their van der Waals surfaces. The picture also shows the location of two crystallographically determined waters around this site in molecule A.



6.9.1 Main chain interactions.

The 2.4Å structure has revealed the high content of secondary structure in the phospholipase A2 molecule. A great number of hydrogen bonds occur between mainchain and sidechain atoms and between sidechains. In this case, a hydrogen bond is assigned if the donor to acceptor distance is between 2.2 and 3.4Å. Table 6.9 lists hydrogen bonds between peptide nitrogens and sidechains. Table 6.10 lists hydrogen bonds between peptide oxygens and sidechains. Table 6.11 lists hydrogen bonds between sidechains.

6.9.2 Helix geometry.

In the alpha helix, the polypeptide chain is normally coiled into a helix with 3.6 residues per turn, and hydrogen bonds between the C=O of residue n and the N-H group of residue n+4. This is the most common type of α-helix in globular proteins with 1->5 hydrogen bonds along the helix. Other types of helix are possible, notably the 3_{10}-helix, with 3 residues per turn, and 1->4 hydrogen bonds, and the π-helix, with 4.4 residues per turn, and 1->6 hydrogen bonds.

Essentially the helical geometry of both equine PLA-2 and bovine PLA-2 is identical. Helix A is a regular α-helix with 1-> 5 contacts ending at residue 13. Helix B is rather irregular, with a mix of 1->5 and 1->4 contacts. Helix C is one of the longest helices in the structure and again is mainly a regular α-helix with 1->5 contacts, but at residue 52 there is a smooth transition to the 3_{10}-helix (from residue 52-58). This is identical to the situation found in the bovine enzyme (Dijkstra *et al.*, 1981). Helix D is also an α-helix, although at the beginning of the helix there is a 1->4 contact. Helix E is approximately the same length as helix C but runs in an anti-parallel

Table 6.9 Hydrogen bonds between peptide nitrogens and sidechains in equine PLA-2.

Description	Distance (Å)	
	Molecule A	Molecule B
27 NH--OD2 42	2.6	2.8
28 NH--OD1 42	2.8	2.9
73 NH--OE1 4	3.0	2.3
74 NH--OD1 88	2.8	2.9
89 NH--OE2 92	3.0	3.2
92 NH--OD1 89	2.8	2.7

Table 6.10 Hydrogen bonds between peptide carbonyls and sidechains in equine PLA-2.

Description	Distance (Å)
A	
6 CO--NE2 10	3.1
9 CO--OG 12	2.8
21 CO--OH 111	2.9
35 CO--OH 28	2.3
48 CO--ND1 48	2.8
112 CO--OH 25	2.7
124 CO--NZ 123	3.0
B	
13 CO--OG 16	2.7
35 CO--OH 28	2.3
67 CO--ND2 67	2.4
85 CO--ND2 88	3.2
112 CO--OH 25	3.4

Table 6.11 Hydrogen bonds between sidechains of equine PLA-2.

Description	Distance (Å)
A	
39 OD1--ND2 112	3.1
48 NE2--OD2 99	3.1
73 OH--OD1 99	2.6
52 OH--OD2 99	2.8
89 ND2--OE2 92	3.1
112 ND2--OD1 115	2.9
B	
21 OE1--NZ 116	3.2
39 OD1--ND2 112	2.4
48 NE2--OD2 99	2.6
52 OH--OD2 99	3.0
73 OH--OD1 99	3.1
115 ND2--OD1 39	2.3

direction. It also has similar geometry to helix C, it being mainly α-helix, but towards the end (from residue 103) the α-helix is distorted and the last residues (104 to 108) form a 3_{10}-helix. This is indicated in plate 6.12 for the Helix E of molecule B.

6.9.3 Protein-solvent contacts.

Hydrogen bonding between the protein and the water molecules is summarised in table 6.12 and 6.13. Hydrogen bonds between solvent molecules are shown in table 6.14. It is clear that most of the water molecules make contacts $\leq$ 3.5 Å with the protein. The waters at the amino terminus which have more than one hydrogen bond have a distinctly reduced freedom of mobility, as measured by their B values. However, most of the solvent forms only single hydrogen bonds to the protein. The majority of these are to the main chain carbonyls with very few being identified for sidechains.

6.10 Conclusions

It would appear that the three-dimensional structure of equine phospholipase is very similar to the high resolution structure of the bovine enzyme (Dijkstra *et al.*, 1981). Due to differences in crystal contacts the conformations around three areas in the two monomers are slightly different. It is proposed that the differences in the calcium binding loop between the two monomers are due to the partial occupancy of the calcium ion in the active site. The majority of the calcium loop is similar to the bovine enzyme and this has been attributed to a number of interactions of the loop with other parts of the structure. Waters have been located in the active site, but they are in a different arrangement to those in the bovine structure and have not

Plate 6.12 Stereo picture of the mainchain structure of helix E of equine PLA-2, showing the hydrogen bonding pattern. This picture shows the α-helix becoming a 3_{10}-helix at residue 103. Hydrogen bonds are illustrated by the broken lines.

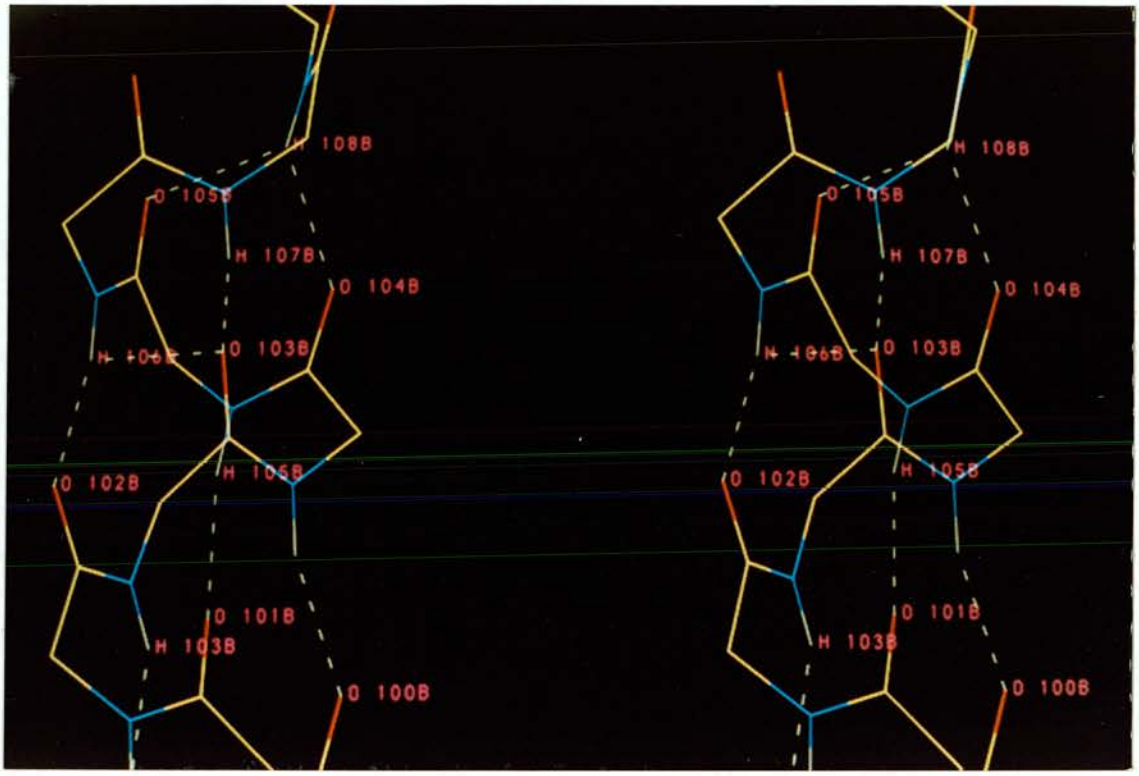


Table 6.12 Hydrogen bonds between protein and solvent.

Description		Distance (Å)	Angle (°)
A	1 NH3+--OH2 1D	2.3	144
	68 CO--OH2 1D	3.1	156
	20 CO--OH2 17D	2.6	157
	44 CO--OH2 11D	2.8	117
	66 NH--OH2 52 D	3.0	133
	72 NH--OH2 20 D	3.2	160
	76 CO--OH2 78 D	3.0	110
	78 NH--OH2 78D	2.9	120
	78 CO--OH2 55D	2.8	141
	82 CO--OH2 70D	3.2	120
	88 CO--OH2 56 D	3.0	123
	107 CO--OH2 26 D	2.4	110
	111 NH--OH2 27D	2.6	167
	113 CO--OH2 58D	3.2	123
	120 NH--OH2 81D	3.2	147
B			
	1 NH3+--OH2 2D	3.1	124
	68 CO--OH2 2D	3.2	129
	31 CO--OH2 6D	2.6	120
	35 CO--OH2 91D	2.5	146
	37 CO--OH2 96D	3.2	105
	39 CO--OH2 96D	3.3	114
	40 CO--OH2 32D	2.5	118
	84 NH--OH2 16D	2.3	151
	84 CO--OH2 99D	2.7	161
	84 CO--OH2 98 D	3.3	109

Table 6.13 Hydrogen bonds between protein sidechains and solvent.

Description	Distance (Å)	Angle (°)
A		
34 OG--OH2 51D	2.9	123
48 ND1--OH2 4D	2.3	99
50 ND2--OH2 46D	3.0	137
53 OG1--OH2 51D	2.6	124
56 NZ--OH2 75D	2.7	158
71 OE1--OH2 8D	3.1	137
71 OE2--OH2 80D	2.6	128
83 OG1--OH2 70D	2.9	167
87 NZ--OH2 44D	3.0	112
88 OD1--OH2 54D	2.8	153
B		
40 OE1--OH2 70D	2.3	144
49 OD1--OH2 5D	2.8	134
108 NZ--OH2 73D	2.9	118
120 OG1--OH2 105D	2.4	107

Table 6.14 Hydrogen bonding between solvent molecules.

Description	Distance (Å)
21D OH2--OH2 70D	2.9
10D OH2--OH2 38D	3.3
52D OH2--OH2 76D	3.1
87D OH2--OH2 88D	2.9
61D OH2--OH2 86D	2.9
30D OH2--OH2 104D	3.3

been assigned as the waters which take part in the catalytic mechanism. The hydrogen bonding network which involves another water molecule is identical to that found in the bovine enzyme. This network is slightly different to that found in the porcine enzyme (Dijkstra *et al.*, 1983). In the equine and bovine structures Pro 68 is involved in the hydrogen bonding of the water at the amino terminus. This may be responsible for the so called '65' loop of the equine enzyme adopting the bovine structure. It was expected that this loop, which has a phenylalanine at position 63 similar to the porcine sequence, would adopt the porcine conformation. In the present structure Asp 66 is now in the correct geometry to be involved in the proposed second calcium binding site along with other sidechains. No calcium ion was found at this site but a number of water molecules were found in this region.

Chapter 7 Conclusions

Chapter 7

Conclusions

During the work of this thesis the crystal structures of two molecules have been determined. The first of these is a small molecule which is an adrenoreceptor antagonist. Certain structural features of this molecule have been proposed as being important in the interaction of this compound with its receptor. This has been proposed on the basis of a number of other molecules with similar biological activity sharing these structural features. Unfortunately these features refer more to the shape of the molecule than to it's specific interaction with the adrenoreceptor. Nonetheless, simple features such as shape are of particular importance in biological interactions. The structural studies on the these small molecules on their own do not provide very much new information to the understanding of their binding to the receptors. It is the hydrogen bonding interactions which are responsible for the specific and tight binding of this molecule to the adrenoreceptor. It is not particularly sensible with so few structures at hand, to attempt to suggest the disposition of donors and acceptors in the binding site of the receptor. However, if there had been some quantitative information available concerning the binding affinities of these related molecules then the study of this structure and the related molecules could have provided better information on the binding. This information is used to produce what is called a quantitative structure-activity relationship (QSAR). Unfortunately this was not available and the only likely solution to the understanding of the interaction of these molecules with their receptor awaits the solution of the receptor structure itself.

It is often not appreciated what a wide range of techniques are required in the production of a crystal structure. These techniques are often nothing in comparison to the beauty of the structure itself and are consequently not given much comment. This is unfortunate but does not mean they were unimportant. Two particular aspects of the work involved in this thesis deserve further mention. The first of which was the purification of the protein in large quantities. Crystallographic investigations require large quantities of absolutely pure protein. This very fact makes the crystallographic investigation of the less abundant proteins either more complicated or unlikely. Chapter 3 described the bulk purification of ovine prophospholipase A2 from pancreas, an abundant source of the enzyme. The purification was made simpler by the protein's ability to survive high temperatures in acidic conditions. It is hoped that the pro-PLA-2 and the activated PLA-2 will be crystallised and the structure of the both forms of the ovine enzyme obtained in the future.

The second technique which lies closer to the solution of the crystal structure itself, is protein crystallisation. This is one of the major stumbling blocks in the process of protein crystallography. So it was with salmon calcitonin (sCT) and its two derivatives, which, despite much effort failed to crystallise. A possible explanation as to the lack of success has been obtained with the discovery that sCT can form amyloid fibres. These fibres are of clinical relevance in that a number of amyloid forming proteins are involved in a whole manner of clinical situations such as Alzheimers' disease. It is highly likely in light of these results, that the gene-related peptide CGRP, also has the ability to form these amyloid fibrils. As mentioned in Chapter 1 CGRP has a high sequence identity with amylin a polypeptide hormone

which has been shown to form amyloid deposits *in vivo*. CGRP as the name suggests has sequence similarities to calcitonin which has been shown in this thesis to also have the potential to form amyloid deposits. Although it could be expected that CGRP could form these fibrous aggregates there has been no clinical evidence to suggest they occur *in vivo*. It is now clear a wide variety of quite diverse proteins have been linked with amyloidosis. Since amyloid forming protein sequences are so diverse, it could not be imagined that one single model of the fibril could be applied to all of them. Thus it has been proposed (Lansbury, 1992) that the defining properties of amyloid characterise a class of related structures, of which the cross-ß fibril (Marsh et al., 1955) is one member rather than the definitive structure. It is possible that the ß-sheet structure identified by NMR in DMSO-water mixtures (Motta et al., 1989) could resemble the conformation in the fibre.

The existing understanding of the action of calcitonin at its receptor is that the active conformation is an α-helix. Thus the formation of these fibres, while explaining the failure of sCT to crystallise, can shed no light on the interaction of sCT with it's receptor. It is evident though, both in light of these results and from various NMR studies, that sCT can adopt a number of quite different conformations depending on the nature of it's environment. Thus the active conformation of sCT can only be determined within some complex. This complex would preferably be with the receptor itself or some soluble extracellular fragment of this receptor. Alternatively the complex could be with a monoclonal antibody which blocks sCT activity - ie. is a mimic of the receptor binding site. Not only would this be easier to crystallise but it would also provide the most

valuable information for the rational design of more active calcitonins.

Even when crystals have been obtained there is the whole process of data collection and processing. This has only been discussed for the crystal structure of the small molecule in Chapter 2. In the crystal structure of equine PLA-2 this work was carried by Dr M. Walkinshaw in Basle. Similarly the process of producing an initial model for a crystal structure, which involves overcoming the phase problem was only mentioned in chapter 2 in relation to small molecules. For protein structures this is much more complicated and has not been discussed in this thesis. Once again this work was carried out for the equine PLA-2 by Dr. M. Walkinshaw. The final stage of the process is the refinement of the initial model. In chapter 5 the procedure of crystallographic refinement was applied in the production of the crystal structure of equine PLA-2. Two methods were successfully applied and the final model possessed a crystallographic residual of 21.7% with data used between 15 and 2.4Å. The refinement of the data obtained from the putative inhibitor-PLA-2 complex failed to indicate the presence of the inhibitor at the active site. This highlights two of the existing problems with inhibitors of PLA-2 which are mimics of the natural substrate. Firstly these types of molecule are very hydrophobic and cannot exist at high concentrations in aqueous environments. Secondly they do not possess sufficient binding affinity to inhibit the enzyme in the presence of the natural substrate which *in vivo* is in vast excess. Both of these problems must be overcome if an effective clinically useful inhibitor of PLA-2 is to be found. The structure of equine PLA-2 has been shown to be very similar to the bovine structure of Dijkstra et al., (1981). A number of slight

differences in the structure were observed, some of which have been explained by the packing interactions in the crystal. However one of the differences has been attributed to the partial occupancy of the active site calcium ion. Since the crystal structure of the equine enzyme has indicated only partial occupancy of the calcium ion in the active site it is not of much use in the drug design process, but it is hoped that further work with this enzyme will lead to novel phospholipase inhibitors in the future.

Bibliography

Bibliography

References

Abita, J-P. and Lazdunski, M. (1972). Eur. J. Biochem. 30, 37-47.

Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. D., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimmer, B. and Sheat, S. (1969). Nature 224, 491-495.

Agarwal, R. C. (1978). Acta Cryst. A34, 791-809.

Alevizaki, M., Shiraishi, A., Rasool, F. V., Ferrier, G. J. M., MacIntyre, I. and Legon, S. (1986). FEBS Lett. 206;47-52.

Allen, F. H., Bellard, S., Price, M.D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. J., Kennard, O., Motherwell, W. D. S., Rodgers, J. R., and Watson, D. G. (1979). Acta Cryst. B 35, 2331-2339.

Arakawa, T. and Timasheff, S. N. (1982). Methods in Enzymology, Vol. 114 pp 49-77.

Axelrod, J., Burch, R. M., and Jelsema, C. L. (1988). Trends Neurochem. Sci. 11, 117-123.

Baghdiantz, A., Foster, G. V., Edwards, A., Kumar, M. A., Slack, E., Soliman, H. A. and MacIntyre, I. (1964). Nature 203, 1027-1028.

Bairaktari, E., Mierke, D. F., Mammi, S., and Peggion, E. (1990). J. Am. Chem. Soc. 112, 5383.

Baker, E. N., and Hubbard, R. E. (1984). Prog. Biophys. Mol. Biol. 44, 97-179.

Baldwin, E. T., Crumley, K. V. and Carter, C. W., Jr. (1986). Biophys. J. 49, 47-48.

Bell, J. A., Moffat, K., Vonderhaar, B. K. and Golde, D. W. (1985). J. Biol. Chem. 260, 8520-8525.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. and Haak, J. R. (1984). J. Chem. Phys. 81(8) 3684-3690.

Betsholtz, C., Johnson, K. H., and Westermark, P. (1989). Nature 338;211.

Blundell, T. L., Pitts, J. E., Tickle, I. J., Woods, S. P. and Wu, C. W. (1981), Proc. Natl. Acad. Sci. U.S.A. 78, 4175-4179.

Bomolaski, J. S., Baker, D. G., Brophy, L. M., and Clark, M. A. (1990). J. Immunology 145, 3391-3397.

Bonar, L., Cohen, A. S., and Skinner, M. M. (1969). Proc. Exp. Biol. Med. 131, 1373-1375.

Bonsen, P. P. M., de Haas, G. H., Pieterson, W. A., and van Deenen, L. L. M. (1972). Biochim. Biophys. Acta. 270, 364-382.

Brain, S. D., MacIntyre, I. and Williams, T. J. (1986). Eur. J. Pharmacol. 124;349-352.

Breimar, L. H., MacIntyre, I. and Zaidi, M. (1988). Biochem. J. 255;377-390.

Brewer, H. B. and Ronan, R.(1969). Proc. Natl. Acad. Sci. U.S.A. 63, 940-942.

Briggs, M. S., Cornell, D. G., Dluhy, R. A. and Gierash, L. M. (1986). Science 233, 206-208.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J. Swaminathan, S. and Karplus, M. (1983). J. Comput. Chem. 4, 187-217.

Brunie, S., Bolin, J., Gewirth, D., and Sigler, P. B. (1985). J. Biol. Chem. 260, 9742-9749.

Brunger, A. T. (1988a). J. Mol. Biol. 203, 803-816.

Brunger, A. T. (1988b). X-PLOR Manual, Version 1.5, Yale

University.

Brunger, A. T., Kuriyan, J. and Karplus, M. (1987). Science 235, 458-460.

Burdick, D., Soreghan, B., Kwon, M., Kosmoski, J., Knauer, M., Henschen, A., Yates, J., Cotman, C., and Glabe, C. (1992). J. Biol. Chem. 267, 546-554.

Burke, J. G. (1966). Origins of the Science of Crystals. University of California Press.

Burns, R. A., and Roberts, M. F. (1980). Biochemistry 19, 3100-3106.

Carter, C. W., Jr., Baldwin, E. T., and Frick, L. (1988). J. Cryst. Growth 90, 60-73.

Carter, C. W., Jr. and Carter, C. W. (1979). J. Biol. Chem. 254, 12219-12223.

Castano, E. M., and Frangione, B. (1988). Lab. Invest. 58, 122-132.

Chausmer, A., Stevens, M. D. and Severn, C. (1982). Science 216;735-736.

Chausmer, A. B., Chavez, C., Wain, R. M. and Daaka, Y. (1989). Metabolism 38, 714-717.

Chiba, T., Yanaguchi, A., Yamatani, T., Nakamura, A., Morishita, T., Inui, T., Fukase, M., Noda, T. and Fujita, T. (1989). Am. J. Physiol. 256;E331-335.

Chou, P. Y. and Fasman, G. D. (1977). J. Mol. Biol. 115, 135-175.

Clark, A., Mathews, D. R. and Naylor, B. A. (1987). Diabetes Res. 4;51-55.

Cochran, M., Peacock, M., Sachs, C. and Nordin, B. E. C. (1970). Brit. Med. J., 1, 135.

Cohen, A. L. (1979). Critical point drying principles and procedures; Scanning electron microscopy 1979, II pp 303-323.

Cohen, A. S., and Skinner, M. (1990). N. Engl. J. Med. 323, 542-545.

Cooper, G. J. S., Willis, A. C., Clark, A., Turner, R. C., Sim, R. B., and Reid, K. B. M. (1987). Proc. Natl. Acad. Sci. U.S.A. 84;8628-8632.

Corey, E. J., Niwa, H., Flack, J. R., Miokowski, C., Arai, Y., and Marfat, A. (1980). Adv. Prostaglandin Thromboxane Res. 6, 19-23.

Cox, M. J. & Weber, P. C. (1987). J. Appl. Cryst. 20, 366-373.

Crews, F. T., Morita, Y., McGivney, A., Hirata, F., Siraganian, R. P. and Axelrod, J. (1981). Arch. Biochim. Biophys. 212;561-571.

Davidson, F. F., Dennis, E. A., Powell, M., and Glenney, J. R., Jr. (1987). J. Biol. Chem. 262, 1698-1705.

Davidson, F. F., Hajdu, J., Dennis, E. A. (1986). Biochem. Biophys. Res. Commun. 137, 587-592.

de Grado,, W. F., Kezdy, F. J., and Kaiser, E. T. (1981). J. Am Chem. Soc. 103, 679-681.

de Haas, G. H., Bonsen, P. P. M., Pieterson, W. A. and van Deenen, L. L. M. (1971). Biochim. Biophys. Acta. 239, 252-266.

de Haas, G. H., Postema, N. M., Nieuwenhuizen, W. and van Deenen, L. L. M. (1968a). Biochim. Biophys. Acta, 159 103-117.

de Haas, G. H., Postema, N. M., Nieuwenhuizen, W. and van Deenen, L. L. M. (1968b). Biochim. Biophys. Acta, 159, 118-129.

Dennis, E. A. (1991). In Phospholipases: Methods in Enzymology Vol. 197, 1-640. Dennis, E. A. ed. Academic Press 1991.

Dijkstra, B. W., Kalk, K. H., Hol, W. G. J., and Drenth, J. (1981a). J. Mol. Biol. 147, 97-123.

Dijkstra, B. W., Drenth, J., and Kalk, K. H. (1981b). Nature 289, 604-606.

Dijkstra, B. W., Renetseder, R., Kalk, K. H., Hol, W. G. J., and Drenth, J. (1983). J. Mol. Biol. 168, 163-179.

Drenth, J., Dijkstra, B. W., and Renetseder, R. (1987). From: Biological Macromolecules and Assemblies; Volume 3-Active Sites of Enzymes. Published by J. Wiley & Sons, Inc. New York, pp. 287-312.

Dufton, M. J., and Hider, R. C. (1983). Eur. J. Biochem. 137, 545-551.

Dutihl, C. E., van Doren, P. J., Verheul, F. E. A. M. and de Haas, G. H. (1975). Eur. J. Biochem., 53, 91-97.

Eanes, E. D. and Glenner, G. G. (1968). J. Histochem. Cytochem. 16, 673-677.

Echlin, P. (1975). Sputter coating techniques for scanning electron microscopy; Scanning electron microscopy 1975, II pp 217-224.

Edsell, J. T. and Wyman, J. (1958). "Biophysical Chemistry" Vol. 1, p 282. Academic Press, New York.

Epand, R. M., Epand, R. F., Orlowski, R. C., Schleuter, R. J., Boni, L.T. and Hui, S. W. (1983). Biochemistry 22, 5074-5084.

Epand, R. M., Epand, R. F. and Orlowski, R. C. (1985). Int. J. Peptide Protein Res. 25, 105-111.

Epand, R. M., Stahl, G. L. and Orlowski, R. C. (1986a). Int. J. Peptide Protein Res. 27, 501-507.

Epand, R. M., Epand, R. F., Orlowski, R. C., Seyler, J. K. and Colescott, R. L. (1986b). Biochemistry 25, 1964-1968.

Epand, R. M., Seyler, J. K. and Orlowski, R. C. (1986c). Eur. J. Biochem. 159, 125-127.

Epand, R. M., Epand, R. F. and Orlowski, R.C. (1988). Biochem. Biophys. Res. Commun. 152, 203-207.

Epand, R. M., Surewicz, W. K., Hughes, D. W., Mantsch, H., Segrest, J. P., Allen, T. M., and Anantharamaiah, G. M. (1989). J. Biol. Chem. 264. 4628-4635.

Evenberg A, Meyer H, Verheij H. M., de Haas, G. H. (1977). Biochim. Biophys. Acta, 491:265-274.

Fehlhammer, H., Bode, W., and Huber, R. (1977). J. Mol. Biol. 111, 415-438.

Figarella C, Clemente F, Guy O. (1971). Biochim. Biophys. Acta, 227:213-217

Findlay, D. M., Ng, K. W., Niall, M. and Martin, T. J. (1982). Biochem. J. 206;343-350.

Fitzgerald, P. M. D. (1988). Acta Cryst. 21, 273-278.

Fletcher, R., and Reeves, C. M. (1965). Comp. J. vol. 7 149-153.

Flower, R. (1989). Biochem. Soc. Trans. 17, 276-278.

Fowlis, W. W., DeLucas, L. J., Twigg, P. J., Howard, S. B., Meehan, E. J. and Baird, J. K. (1988). J. Cryst. Growth 90, 117-129.

Fraser, P. E., Nguyen, J. T., Surewicz, W. K., and Kirschner, D. A. (1991). Biophys. J. 60, 1190-1201.

Fritz, H., Huller, I., Wiedemann, M. and Werle, E. (1967). Z. Physiol. Chem., 348 405.

Galente, L., Horton, R., Joplin, G. F., Woodhouse, N. J. F. and MacIntyre, I. (1971). Clin. Sci. 40, 9-10.

Galente, L., Joplin, G. F., MacIntyre, I. and Woodhouse, N. J. Y.

(1976). Clin. Sci. Mol. Med. 44;605-610.

Gennari, C., Chierichetti, S. M., Vibelli, C., Francini, G.,Maioli, E. (1981). Curr. Ther. Res. 30;1024-1032.

Gibson, S. J., Polak, J. M., Bloom, S. R., Sabate, I. M., Muldberry, P. K., Ghatei, M. A., McGregor, J. P., Morrison, J. F. B., Kelly, J. S., Evans, R. M. and Rosenfeld, M. G. (1984). J. Neurosci. 4;1301-1311.

Glaser, K. B., and Jacobs, R. S. (1986). Biochem. Pharmac. 35, 449-453.

Glenner, G. G. (1980a). N. Engl. J. Med. 302, 1283-1292.

Glenner, G. G. (1980b). N. Engl. J. Med. 302, 1333-1343.

Glennon, R.A. (1987). Journal of Medicinal Chemistry, 30, 1.

Glennon, R.A., Naiman, N.A., Lyon, R.A. & Titeler, M. (1988). Journal of Medicinal Chemistry, 31, 1968-1971.

Goltzmann, D., and Mitchell, J. (1985). Science (Washington D.C.) 227;1343-1345.

Gould, R.O. & Taylor, P. (1983). CALC. Interactive program for molecular geometry, University of Edinburgh, Scotland.

Green, F. R., Lynch, B. and Kaiser, E. T. (1987). Proc. Natl. Acad. Sci. 84, 8340-8344.

Guttmann, S. (1981). In:Calcitonin 1980. Chemistry, Physiology, Pharmacology and Clinical Aspects, Pecile, A. (Editor), pp. 11-24, Exerpta Medica, Amsterdam.

Halverson, K., Fraser, P. E., Kirschner, D. A., and Landsbury, Jr. P. T. (1990). Biochemistry 29, 2639-2644.

Hanahan, D. J. (1962). J. Biol. Chem., 195 199.

Hayaishi, O. and Kornberg, A. (1954). J. Biol. Chem. 206, 647.

Heinrikson, R. L., Kreuger, E. T., and Keim, P. S. (1977). J.

Biol. Chem. 252, 4913-4921.

Hershberg, R. D., Reed, G. H., Slotbloom, A. J., and de Haas, G. H. (1976). Biochemistry 15, 2268-2274.

Hogle, J. M., Chow, M., and Filman, D. J. (1985). Science 229,1358-1365.

Hoppener, J. W. M., Steenburgh, P. A., Zandbergh, J., Geurts van Kessel, A. H. M., Baylin, S. B., Nelkin, B. D., Jansz, H. S. and Lips, C. J. M. (1985). Hum. Genet. 70;259-263.

Howard, B. D., and Gundersen, C. B., Jr. (1980). Ann. Rev. Pharmacol. Toxicol. 20, 307-336.

Howard, C. F. (1986). Diabetologia 29;301-306.

Hubbard, J. A. M., Martin, S. R., Chaplin, L. C., Bose, C., Kelly, S. M., and Price, N. C. (1991). Biochem. J. 275;785-788.

International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham, England: Kynoch Press.

IUPAC-IUB Commission on Biochemical Nomenclature (1970). J. Mol. Biol. 52, 1-17.

Jack, A. and Levitt, M. (1978). Acta Cryst. A34, 931-935.

Jhoti, H. (1989). In: Molecular simulation and protein crystallography pp 42-53. Proceedings of the joint CCP4/CCP5 study weekend. S.E.R.C. Daresbury.

Johnson, C.K. (1965). ORTEP. Report ORNL-3794. Oak Ridge National Laboratory, Tenessee, U.S.A.

Johnson, K. H. and Stevens, J. B. (1973). Diabetes 22;81-90.

Johnson, K. H. (1987). Proc. Natl. Acad. Sci. U.S.A. 84;3881-3885.

Jonas, V., Lin, C. R., Kawashima, E., Semon, D., Swanson, L. W., Mermod, J.-J., Evans, R. M. and Rosenfeld, M. G. (1985).

Proc. Natl. Acad. Sci. 82;1994-1998.

Jones, A. (1978). FRODO. A graphics model building and refinement system for macromolecules. J. Appl. Cryst. 11, 268-272.

Jones, A. J. S., Epand, R. M., Lin, K. F., Walton, D. and Vail, W. J. (1978). Biochemistry 17, 2301.

Kabsch, W., and Sander, C. (1983). Biopolymers 22, 2577-2637.

Kaiser, E. T., and Kezdy, F. J. (1983). Proc. Natl. Acad. Sci. 80, 1137-1143.

Kam, Z., Shore, H. B. and Feher, G. (1987). J. Mol. Biol. 123, 539-555.

Karle, I. L., and Duesler, E. (1977). Proc. Natl. Acad. Sci., U.S.A. 74, 2602-2606.

Karlsson, E. (1978). Chemistry of protein toxins in snake venoms. In: Handbook for Experimental Pharmacology, Vol. 52, Snake Venoms, 159-212 (Lee, C. Y., Ed.). Berlin: Springer.

Kauzmann, W. (1959). Adv. Protein Chem. 14, 1.

Kirkpatrick, S. Gelatt, C. D. and Vecchi, M. P. (1983). Science 220, 671-680.

Kirschner, D. A., Inouye, H., Duffy, L. K., Sinclair, A., Lind, M., and Selkoe, D. J. (1987). Proc. Natl. Acad. Sci. 84, 6953-6957.

Klee, C. B. (1988). Biochemistry 27, 6645-6653.

Kreusch, A., Weiss, M. S., Welte, W., Weckesser, J., and Schulz, G. E. (1991). J. Mol. Biol. 217, 9-10.

Kubota, M., Moseley, J. M., Butera, L., Dusting, G. J., MacDonald, P. S. and Martin, T. J. (1985). Biochem. Biophys. Res. Commun. 138;88-94.

Laemmli, U. K. (1970). Nature 227, 680-685.

Lansbury, Jr., P. T. (1992). Biochemistry 31, 6865-6870.

Leffert, J. D., Newgard, C. B., Okamato, H., Milburn, J. L. and

Leger, J.-M., Saux, M., and Carpy, A (1983) Acta Cryst. C39, 1428-1430.

Leger, J.-M., Colleter, J.-C., and Carpy, A. (1983). Eur. J. Med. Chem.-Chim. Ther. 18, 559-561.

Levitt, M. (1982). Ann Rev. Biophys. Bioeng. 11, 251-271.

Lichtenberg, D., Romero, G., Menashe, M., and Biltonen, R. L. (1986). J. Biol. Chem. 261, 5334-5340.

Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951). J. Biol. Chem. 193, 265-275.

Luskey, K. L. (1989). Proc. Natl. Acad. Sci. U.S.A. 86; 3127-3130.

Luben, R. A., Wong, G. L. and Cohn, D. V. (1976). Endocrinology 99;526-534.

Lynch, B. and Kaiser, E. T. (1988). Biochemistry 27;7600-7607.

MacIntyre, I., Alevizaki, M., Bevis, P. J. R., and Zaidi, M. (1987). Clin. Ortop. Relat. Res. 21, 45-54.

Magee, W. L., Gallai-Hatchard, J., Saunders, H. and Thompson, R. H. S. (1962). Biochem. J., 83 17-25.

Maier, R., Brugger, M., Bruckner, H., Kamber, B., Riniker, B. and Rittel, W. (1977). Acta. Endocrinol. (Copenhagen) 85, 102.

Marsh, R. E., Corey, R. B., and Pauling, L. (1955). Biochim. Biophys. Acta 16, 1-34.

Marx, S. J., Woodward, C. J. and Aurbach, G. D. (1972). Science 178;998-1001.

Marx, S. J., Aurbach, G. D., Gavin, J. R. and Guell, J. W. (1974). J. Biol. Chem. 249;6812-6816.

McGivney, A., Morita, Y., Crews, F. T., Hirata, F., Axelrod, J. and Siraganian, R. P. (1981). Arch. Biochim. Biophys. 212;572-580.

Mc Kie, D., and Mc Kie, C. (1986). Essentials of crystallography. Blackwell scientific publications, Oxford.

McPherson, A. (1982). Preparation and Analysis of Protein Crystals, Wiley, New York.

Meadows, R. P., Nikonowicz, E. P., Jones, C. R., Bastian, J. W., and Gorenstein, D. G. (1991). Biochemistry 30, 1247-1254.

Meyer, J.-P., Pelton, J. T., Hoflack, J. and Saudek, V. (1991). Biopolymers 31, 233-241.

Mikol, V., Hirsch, E. and Giege, R. (1990a). J. Mol. Biol. 213, 187-195.

Mikol, V., Rodeau, J. -L. and Giege, R. (1990b). Analytical Biochemistry 186, 332-339.

Moe, G. R. and Kaiser, E. T. (1985). Biochemistry 24, 1971-1976.

Morikawa, T., Munekata, E., Sakakibara, S., Noda, T. and Otani, M. (1976). Experentia 32, 1104-1106.

Morris, H. R., Panico, M., Etienne, T., Tippins, J., Girgis, S. I. and MacIntyre, I. (1984). Nature 308;746-748.

Motherwell, W.D.S. (1972). PLUTO. Program for plotting molecular and crystal structures. University of Cambridge, England.

Motta, A., Antonietta, M., Morelli, C., Goud, N. and Temussi, P. A. (1989). Biochemistry 28, 7996-8002.

Nicholson, G. C., Moseley, J. M., Sexton, P. M., Mendelsohn, P. A. O. and Martin, T. J. (1986). J. Clin. Invest. 78;355-360.

Nieuwenhuizen, W., Steenbergh, P. and de Haas, G. H. (1973). Eur. J. Biochem 40, 1-7.

Nieuwenhuizen, W., Kunze, H. and de Haas, G. H. (1973). Methods in Enzymology Vol 32B:147-154.

Nishi, M., Sanke, T., Nagamatsu, S., Bell, G. I., and Steiner, D. F. (1990). J. Biol. Chem. 265, 4173-4176.

O'Dor, R. K., Parkes, C. D. and Copp, D. H. (1969). Can. J. Biol. 47, 873-875.

Osborne, T. B. (1892). Amer. Chem. J. (Baltimore) 14, 662.

O'Riordan, J. H. and Auerbach, G. D. (1968). Endocrinology 82, 377-380.

Orlowski, R. C., Epand, R. M. and Stafford, A. R. (1987). Eur. J. Biochem. 162, 399-402.

Otani, M., Yamauchi, H., Meguro, T., Kitazawa S., Wanatabe, S. and Orimo, H. (1976). J. Biochem. (Tokyo) 79, 345-352.

Pepys, M. B. (1988). Amyloidosis. In Immunological disease (4th edition ed. M. Samter et al.). Little, Brown and Co., Boston.

Pieterson, W. A., Vidal, J. C., Volwerk, J. J., and de Haas, G. H. (1974a). Biochemistry 13, 1455-1460.

Pieterson, W. A., Volwerk, J. J., and de Haas, G. H. (1974b). Biochemistry, 13, 1439-1445.

Potts, J. T., Jr., Niall, H. D., Keutman, H.T., Deftos, L. J. and Parsons, J. A. (1970) in Calcitonin : Recent Chemical and Immunological studies. Calcitonin 1969: Proceedings of Second International Symposium (Taylor, S. ed.), pp 56-73, Heinmann Medical Books, London.

Pruzanski, W., and Vadas, P. (1988). J. Rheumatology 15, 1601-1603.

Puchtler, H. Sweat, F. and Levine, M. (1962). J. Histochem. Cytochem. 10, 355-364.

Pullman, B., Coubeils, J.-L., Courrriere, P. H., and Gervois, J.-P. (1972). J. Med. Chem. 15, 17-23.

Raulias, D. Hagman, J., Ontjes, D. A., Lundblad, R. L. and Kingdon, H. S. (1976). Eur. J. Biochem. 64, 607-611.

Reynolds, L. J., Morgan, B. P., Hite, G. A., Mihelich, E. D., and Dennis, E. A. (1988). J. Am. Chem. Soc. 110, 5172-5177.

Richard, M. (1989). J. Rheumatology (suppl 18) 16, 35-38.

Rimon, A. and Shapiro, B. (1959). Biochem. J., 71 620-623.

Riniker, B., Neher, R., Maier, R., Kahnt, F. W., Byfield, P. G. H., Gudmundsson, T. V., Galente, L. and MacIntyre, I. (1968) Helv. Chim. Acta. 51, 1738-1742.

Rittel, W., Maier, R., Brugger, M., Kamber, B., Riniker, B. and Sieber, P. (1976). Experentia 32, 246-248.

Roberts, M. F., Deems, R. A., Mincey, T. C., and Dennis, E. A. (1977). J. Biol. Chem. 252, 2405-2411.

Roberts, M. F., Bothner-By, A. A., and Dennis, E. A. (1978). Biochemistry 17, 935-941.

Rosenfeld, M. G., Mermod, J.-J., Amara, S. G., Swanson, L. W., Sawchenko, P.E., Rivier, J., Vale, W. W. and Evans, R. M. (1983). Nature 304;129-135.

Saldanha, J. and Mahadevan, D. (1991). Protein Engineering vol.4 no.5;539-544.

Sasaki, K., Dockerill, S., Adamiak, D. A., Tickle, I. J., and Blundell, T. (1975). Nature 257, 751-758.

Scott, D. L., Otwinowski, Z., Gelb, M. H., and Sigler, P. B. (1990). Science 250, 1563-1566.

Segrest, J. P., Jackson, R. L., Morrisett, J. D. and Gotto, A. M., Jr., (1974). FEBS Lett. 38, 247.

Selkoe, D. J. (1991). Neuron 6, 487-498.

Sheldrick, G.M. (1986). SHELXS86. Program for crystal structure solution University of Gottingen, Federal Republic of Germany.

Sletten, K., Westermark, P., and Natvig, J. B. (1976). J. Exp. Med. 143, 993-998.

Slotbloom, A. J., Jansen, E. H. J. M., Vlijm, H., Patuus, F., de Araujo, P. S., and de Haas, G. H. (1978). Biochemistry 17, 4593-4600.

Spencer, R. G. S., Halverson, K. J., Auger, M, McDermott, A. E., Griffin, R. G., and Lansbury, Jr., P. T. (1991). Biochemistry 30, 10382-10387.

Stout, G.H. & Jensen, L.H. (1969). X-ray Structure Determination A Practical Guide. Collier Macmillan Publishers, London.

Strikland-Constable, R. F. (1968). Kinetics and Mechanism of Crystallisation, Academic Press, New York.

Ten Eyck, L. F. (1977). Acta Cryst. A33, 486-492.

Terwilliger, T. C., Weissman, L., and Eisenberg, D. (1982a). J. Biol. Chem. 257, 6010-6015.

Terwilliger, T., Weissman, L. and Eisenberg, D. (1982b). Biophys. J. 37, 353-357.

Thornton, J. M. (1981). J. Mol. Biol. 151, 261-287.

Thunnissen, M. M. G. M., Kalk, K. H.,Drenth, J., and Dijkstra, B. W. (1990a). J. Mol. Biol. 216, 425-439.

Thunnissen, M. M. G. M., Ab, E., Kalk, K. H., Drenth, J., Dijkstra, B. W., Kuipers, O. P., Dijkman, R., de Haas, G. H., and Verheij, H. M. (1990b). Nature 347, 689-691.

Tippins, J. R., Morris, H. R., Panico, M., Etienne, T., Bevis, P.,

Girgis, S., MacIntyre, I., Azria, M. and Attinger, M. (1984). Neuropeptides 4;425-434.

Treece, J. M., Sheinson, R. S. and Mc Meekin, T. L. (1964). Arch. Biochem. Biophys. 108, 99.

Tripos Associates, SYBYL: A molecular modelling package. Tripos Associates Inc., St Louis Missouri (1991).

Tronrud, D. E., Ten Eyck, L. F. and Matthews, B. W. (1987). Acta Cryst. A43, 489-501.

Tschopp, F. A., Henke, H., Petermann, J. B., Tobler, P. H., Janzer, R., Hokfelt, T., Lundberg, J. M., Cuello, C. and Fisher, J. A. (1985). Proc. Natl. Acad. Sci. U.S.A. 82;248-252.

Vadas, P., and Pruzanski, W. (1986). Lab. Invest. 55, 397-404.

van Dam-Mieras, M. C. E., Slotbloom, A. J., Pieterson, W. A., and de Haas, G. H. (1975). Biochemistry 14, 5387-5394.

van den Bergh, C. J., Bekkers, A. C. A. P. A., Verheij, H. M., and de Haas, G. H. (1989). Eur. J. Biochem. 182, 307-313.

van den Bosch, H. (1980). Biochimica et Biophysica Acta, 604;191-246.

van den Bosch, H., Postema, N. M., de Haas, G. H. and Heemskerk, C. H. T. (1965). Biochim. Biophys. Acta, 98 657.

van Kuijk, F. J. G. M., Sevanian, A., Handelman, G. J., and Dratz, E. A. (1987). Trends Biochem. Sci. 12, 31-34.

Verger, R., and de Haas, G. H. (1976). Annu Rev. Biophys. Bioeng. 5, 77-119.

Verheij H. M., Slotboom, A. J. and de Haas G. H. (1981). Rev. Physiol. Biochem. Pharmacol., Vol 91:91-227.

Verheij, H. M., Volwerk, J. J., Jansen, E. H., Puyck, W. C., Dijkstra, B. W., Drenth, J., de Haas, G. H. (1980). Biochemistry

19, 743-750.

Vignon, E., Mathieu, P., Louisot, P., Vilamitjana, J., Harmand, M. F., and Richard, M. (1989). J. Rheumatol. 16(suppl 18), 35-38.

Vogt, W., Suzuli, T., and Babilli, S. (1966). Mem. Soc. Endocrinol. 14, 137-142.

Volwerk, J. J., and de Haas, G. H. (1982). In: Lipid-Protein Interactions, P. C. Jost, and O. H. Griffiths, Eds., Vol. 1. J. Wiley & Sons. New York, pp. 69-149.

Volwerk, J. J., Pieterson, W. A., and de Haas, G. H. (1974). Biochemistry 13, 1446-1454.

Warshawsky, F., Goltzmann, D., Rouleau, M. F., Andergeron, J. M. (1980). J. Cell Biol. 88;682-694.

Waszkowycz, B., Hillier, I. H., Gensmantel, N., and Payling, D. W. (1989). J. Chem. Soc. Perkin Trans. II, 1795-1800.

Wery, J.-P., Schewitz, R. W., Clawson, D. K., Bobitt, J. L., Dow, E. R., Gamboa, G., Goodson, Jr., T., Hermann, R. B., Kramer, R. M., McClure, D. B., Mihelich, E. D., Putnam, J. E., Sharp, J. D., Stark, D. H., Teater, C., Warrick, M. W., and Jones, N. D. (1991). Nature 352, 79-82.

Westermark, P., Grimelius, L., Polak, J. M.,Larsson, L.-I., Van Noorden, S.,Wilander, E., and Everson Pearse, A. G. (1977). Lab. Invest. 37, 212-215.

Westermark, P., Wernstedt, C., Wilander, E., Hayden, D. W., O'Brien, T. D. and Steenbergh, P. H., Hoppener, J. W. M., Zandberg, J., Lips, C. J. M. and Jansz, H. S. (1985). FEBS Lett. 183;403-407.

White, S. P., Scott, D. L., Otwinowski, Z., Gelb, M. H., and Sigler P. B. (1990). Science 250, 1560-1563.

Wood, S. P., Pitts, J. E., Blundell, T. L., Tickle, I. J. and Jenkins, J. A. (1977). Eur. J. Biochem. 78, 119-126.

(1991). J. Mol. Biol. 217, 9-10.

Wood, R. H. and Thompson, P.T. (1990). Proc. Natl. Acad. Sci. 87, 946 - 949.

Woolfson, M.M. (1970). An Introduction to X-ray Crystallography. Cambridge University Press, England.

Wüthrich, K. (1976). In "NMR in Biological Research: Peptides and Proteins", pp 65-72, North Holland, Amsterdam.

Yamamoto, I., Morita, R., Fukunaga, M., Dokoh, S., Shigeno, C., Torikuza, K. and Noda, T. (1981). Endocrinology 108;698-702.

Zaidi, M., Chambers, T. J., Gaines Das, R. E., Morris, H. R. and MacIntyre, I. (1987). J. Endocrinol. 115;511-518.

Zaidi, M., Brain, S. D., Tippins, J. R., Di Marzo, V., Moonga, B. S., Chambers, T. J., Morris, H. R. and MacIntyre, I. (1990). Biochem. J. 269;775-780.